Estimating a dose-response function with heterogeneous response to confounders when treatment is continuous and endogenous

Christopher F. Baum
Department of Economics
Boston College
E-mail: baum@bc.edu

Giovanni Cerulli
CNR-IRCrES
National Research Council of Italy
Institute for Research on Sustainable Economic Growth
E-mail: giovanni.cerulli@ircres.cnr.it

Structured Abstract

Background: This paper presents an original econometric model for estimating a dose-response function when: treatment is continuous but with a "spike" at zero; individuals may react heterogeneously to observable confounders; and selection-into-treatment may be endogenous.

Objective: The goal is to provide program evaluators with a model able to enlarge the set of program evaluation methods to use in continuous treatment settings.

Methodology: Two estimation procedures are suggested: OLS under Conditional Mean Independence; and Instrumental-Variables (IV) under treatment endogeneity.

Empirical illustration: Two applications to real data are proposed. The first investigates the effect of public R&D support on companies' R&D expenditure for Italy using the OLS procedure; the second application uses US data to assess the effect of job tenure on wages. Results in the light of model's assumptions are discussed.

Keywords: treatment effects, dose-response function, continuous treatment

JEL classification: C21, C87, D04

25 January 2016

1. Introduction

In many socio-economic contexts, policy interventions take the form of a 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 typical examples in which the amount of support can vary by individual, thereby providing ground for a different response to policy.

Therefore, 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") provided by a public agency. This is in tune with the language of epidemiology, where dose-response functions are usually estimated in order to check patients' resilience to different levels of drug administration (Robertson et al., 1994; Royston and Sauerbrei, 2008).

This paper presents an original econometric model for estimating a dose-response function through a regression approach when: (i) treatment is continuous, (ii) individuals may react heterogeneously to observable confounders, and (iii) selection-into-treatment may be potentially endogenous.

To fix ideas, consider a policy program where the treatment is not randomly assigned (i.e., assigned according to some "structural" rule), and where – after setting who is treated and who is not – the program provides a different "level" or "exposure" to treatment ranging from 0 (no 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* based on the observed sample, by assuming that treated and untreated units may respond differently both to specific observable confounders (that we collect in a row vector **x**), and to the "intensity" of the treatment *t*. We wish to estimate a dose-response function of *y* on *t*, either when the treatment is assumed to be exogenous (i.e., selection-into-treatment depends only on observable-to-analyst factors) or endogenous (i.e., selection-into-treatment depends both on observable and unobservable-to-analyst factors).

In this model, the dose-response function is shown to be equal to the "Average Treatment Effect, given the level of treatment *t*" (i.e. ATE(*t*)). But also other causal parameters of interest, such as the unconditional Average Treatment Effect (ATE), the Average Treatment Effect on Treated (ATET), the Average Treatment Effect on Non-Treated (ATENT) are estimated, along with these effects conditional on the vector (**x**; *t*).

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 (the so-called "spike at zero"). Additionally, it may account for treatment endogeneity by exploiting an Instrumental-Variables (IV) estimation in a continuous treatment context (provided that good instruments are available).

When many units are not exposed to treatment, the distribution of *t* has a "spike" or non-nil probability mass at zero, i.e. Pr(*t*=0)>0. This means that assuming that the

[1] See also Bia et al. (2014) generalizing the Hirano-Imbens (2004) model by allowing for a nonparametric estimation of the Dose-Response Function.

distribution of $t|\mathbf{x}$ comes from a normal (or mixtures of normal) distribution, as assumed in the Generalized Propensity Score (GPS) proposed by Hirano and Imbens, is untenable, as in the presence of a spike at zero this distribution is clearly discontinuous and thus non normal. Recently, however, Guardabascio and Ventura (2014) have proposed a generalization of the Hirano and Imbens model extending the GPS approach to the case of a non-normal continuous treatment variable. They consider a set of alternative distributions (binomial, poisson, gamma, inverse-gaussian, etc.) derived from the exponential family distribution. Although rich in its scope, such a model is still unsuited to incorporate zero-treatment and potential treatment endogeneity. The present paper tries to overcome both these limitations.

Within the epidemiological literature, Royston et al. (2010) have proposed a dose-response model for continuous exposures with a spike at zero based on fractional polynomials. For fractional polynomials functions of t to be defined at t=0, the authors shift the origin of t by adding a small constant, c, before analysis. They take c as the smallest difference between successive observed positive values of t, although other choices are suggested. The authors propose a model for the response variable y having a jump in zero with y equal to a constant β in t=0 and to a fractional polynomial of $t+c$ in $t>0$. They estimate this model in a single regression by adding an indicator variable z (taking value one if t=0 and zero if $t>0$) as additional predictor. In this way they are able to estimate the response at t=0 (i.e., recovering a consistent estimation of β) by exploiting a standard regression model.

Differently from Royston et al., our model is embedded into the potential outcome setting, the typical framework of counterfactual modeling, and it solves the

zero-inflation problem by directly modeling the potential outcome in t=0. In this way, we avoid ad hoc assumptions, as that for choosing a reliable constant c.

Differently from Hirano and Imbens, we do not need to specify a generalized propensity score, as we work within a control-function model[2]. Moreover, we are able to take into account both zero-inflation at t=0 and treatment endogeneity under reasonable assumptions. More specifically, we model the dose-response function as approximated by a third degree polynomial and both OLS and IV estimation is considered. IV is based on a Heckman bivariate selection (also known as type-2 tobit) 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 two-stage least squares (2SLS) estimation for the outcome (y) equation in the second step.

The paper presents two illustrative applications of the proposed model to real datasets. The first application performs the OLS approach assuming treatment

[2] As Wooldridge (2010, p. 924-925) points out, in a control-function regression setting one cannot do better than using the entire set of available covariates as controls instead of the propensity score as single control. Using the propensity score as the only regression control induces in fact two additional problems: (1) the propensity score is a rough representation of **x** and is "estimated", thus subject to sampling error; this induces some noise, generally leading to larger estimates' standard errors (i.e., lower precision). (2) If the actual data generating process of the propensity score does not follow the probit/logit form, one runs the risk to insert into the regression a variable affected by measurement error, thus leading to inconsistent OLS estimates of the treatment effects. As a conclusion, one does not have any plausible reason to prefer the use of the propensity score instead of the entire set of covariates within a regression setting. Differently, if one uses Matching instead of a regression approach, then the use of the propensity score becomes necessary to help the researcher to overcome the so-called "dimensionality problem" (Rosenbaum and Rubin, 1983; Dehejia and Wahba, 2002), although further complications can arise also in this case (Abadie and Imbens, 2012).

(conditional) exogeneity; the second application applies the IV procedure assuming treatment endogeneity. In the first case, we assess the effect of public R&D support on business R&D outlay. We will show how our model can be able to look at the pattern of the policy effect over treatment intensity, thus going beyond the typical "average effect" analysis. This suitably allows us for a better inspection into the causal relation between the policy instrument and the policy target. In the second application, we try to estimate the impact of job tenure on wages on a dataset of women aged between 14 and 26 years by assuming tenure to be endogeneous, thus relying on the proposed IV approach.

The paper is organized as follows: section 2, 3 and subsections present the model, its assumptions and propositions, as well as the related estimation techniques; section 4 sets out the OLS (section 4.1) and the IV (section 4.2) applications on real data; section 5, finally, concludes the paper. At the end of the paper, appendix A presents the proofs of the propositions.

2. The model

We set out with some notation. Consider two different and exclusive outcomes: one referring to a unit i when she is treated, y_{1i}; and one referring to the same unit when she is untreated, y_{0i}. Define w_i as the treatment indicator, taking value 1 for treated and 0 for untreated units, and $\mathbf{x}_i = (x_{1i}, x_{2i}, x_{3i}, \ldots, x_{Mi})$ as a row vector of M exogenous observable characteristics (confounders) for unit $i = 1, \ldots, N$. Let N be the number of units involved in the experiment, N_1 be the number of treated units, and N_0 the number of untreated units with $N = N_1 + N_0$.

Define two distinct functions, $g_1(\mathbf{x}_i)$ and $g_0(\mathbf{x}_i)$, as the unit i's responses to the vector of confounding variables $\mathbf{x}_i$ 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 having zero unconditional mean and constant variance. Finally, define t_i – taking values within the continuous range [0;100] – as the continuous-treatment indicator, and $h(t_i)$ as a general derivable function of t_i. In what follows, in order to simplify notation, we'll get rid of the subscript i when defining population quantities and relations.

Given previous notation, we assume a specific population generating process for the two exclusive potential outcomes[3].

Assumption 1. *Form of the potential outcomes' population generating process*.

Given the previous definitions, the potential outcomes are modelled in an additive form:

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

where the $h(t)$ function is different from zero only in the treated status:

$$\begin{cases} h(t) = 0 & if \quad w = 0 \\ h(t) \neq 0 & if \quad w = 1 \end{cases} \tag{2}$$

Given previous assumption and notation, we can also define the causal parameters of interests.

[3] Such a model is the representation of a *treatment random coefficient regression* as showed by Wooldridge (1997; 2003). See also Wooldridge (2010, Ch. 18). For the sake of simplicity, as we refer to the population model, here we avoid to write the subscript i referring to each single unit i's relationships.

Definition 1. By defining the treatment effect as TE = ($y_1 - y_0$), we define the causal parameters of interest as the population Average Treatment Effects (ATEs) conditional on **x** and t, that is:

$$\begin{aligned} &\text{ATE}(\mathbf{x};t) = \text{E}(y_1 - y_0 \mid \mathbf{x},t) \\ &\text{ATET}(\mathbf{x};t>0) = \text{E}(y_1 - y_0 \mid \mathbf{x},t>0) \\ &\text{ATENT}(\mathbf{x};t=0) = \text{E}(y_1 - y_0 \mid \mathbf{x},t=0) \end{aligned} \tag{3}$$

where ATE indicated the average treatment effect, ATET the average treatment effect on treated, and ATENT the one on untreated units. By the law of iterated expectation (LIE), we know that the population unconditional ATEs are obtained as:

$$\begin{aligned} &\text{ATE} = \text{E}_{(\mathbf{x};t)}\{\text{ATE}(\mathbf{x};t)\} \\ &\text{ATET} = \text{E}_{(\mathbf{x};t>0)}\{\text{ATE}(\mathbf{x};t>0)\} \\ &\text{ATENT} = \text{E}_{(\mathbf{x};t=0)}\{\text{ATE}(\mathbf{x};t=0)\} \end{aligned} \tag{4}$$

where $\text{E}_{(\mathbf{v})}\{\cdot\}$ identifies the mean operator taken over the support of a generic vector of variables **v**.

By assuming a linear parametric form for $g_0(\mathbf{x}) = \mathbf{x}\boldsymbol{\delta}_0$ and $g_1(\mathbf{x}) = \mathbf{x}\boldsymbol{\delta}_1$ the Average Treatment Effect (ATE) conditional on **x** and t becomes:

$$\begin{aligned} &\text{ATE}(\mathbf{x};t) = \text{E}(y_1 - y_0 \mid \mathbf{x},t) = \begin{cases} (\mu_1 - \mu_0) + \mathbf{x}(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_0) + h(t) & \text{if } t>0 \\ (\mu_1 - \mu_0) + \mathbf{x}(\boldsymbol{\delta}_1 - \boldsymbol{\delta}_0) & \text{if } t=0 \end{cases} = \\ &= \begin{cases} \mu + \mathbf{x}\boldsymbol{\delta} + h(t) & \text{if } t>0 \\ \mu + \mathbf{x}\boldsymbol{\delta} & \text{if } t=0 \end{cases} \end{aligned} \tag{5}$$

or also:

$$\mathrm{ATE}(\mathbf{x},t,w) = w\cdot[\mu + \mathbf{x}\boldsymbol{\delta} + h(t)] + (1-w)\cdot[\mu + \mathbf{x}\boldsymbol{\delta}] \tag{6}$$

where μ=(μ_1-μ_0) and $\boldsymbol{\delta}$=($\boldsymbol{\delta}_1$-$\boldsymbol{\delta}_0$). Thus, we can state this proposition:

Proposition 1. Given previous notation and definitions, the unconditional Average Treatment Effect (ATE) related to model (1) is equal to:

$$\mathrm{ATE} = p(w=1)\cdot(\mu + \overline{\mathbf{x}}_{t>0}\boldsymbol{\delta} + \overline{h}_{t>0}) + p(w=0)\cdot(\mu + \overline{\mathbf{x}}_{t=0}\boldsymbol{\delta})$$

where $p(\cdot)$ is a probability, and $\overline{h}_{t>0}$ is the average of the response function taken over t>0. Appendix A provides the proof. See A1.

Since, by LIE, we have that ATE = $p(w$=1)·ATET + $p(w$=0)·ATENT, we obtain from the previous formula that:

$$\begin{cases} \mathrm{ATE} = p(w=1)(\mu + \overline{\mathbf{x}}_{t>0}\boldsymbol{\delta} + \overline{h}_{t>0}) + p(w=0)(\mu + \overline{\mathbf{x}}_{t=0}\boldsymbol{\delta}) \\ \mathrm{ATET} = \mu + \overline{\mathbf{x}}_{t>0}\boldsymbol{\delta} + \overline{h}_{t>0} \\ \mathrm{ATENT} = \mu + \overline{\mathbf{x}}_{t=0}\boldsymbol{\delta} \end{cases} \tag{7}$$

with $\overline{\mathbf{x}}_{t=0}$ and $\overline{\mathbf{x}}_{t>0}$ equal to the mean of $\mathbf{x}$ in t=0 and over t>0 respectively. By adding and subtracting the same expressions, we then obtain that:

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

leading to:

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

so that:

$$\begin{cases} \mathrm{ATET}(\mathbf{x},t) = \mathrm{ATE}(\mathbf{x},t,w=1) = \mathrm{ATET} + (\mathbf{x}_{t>0} - \overline{\mathbf{x}}_{t>0})\boldsymbol{\delta} + (h(t) - \overline{h}_{t>0}) \\ \mathrm{ATE}(\mathbf{x},t) = \mathrm{ATE}(\mathbf{x},t,w=0) = \mathrm{ATENT} + (\mathbf{x}_{t=0} - \overline{\mathbf{x}}_{t=0})\boldsymbol{\delta} \end{cases} \tag{8}$$

where:

$$\begin{cases} \mathrm{ATET} = \mu + \overline{\mathbf{x}}_{t>0}\boldsymbol{\delta} + \overline{h}_{t>0} \\ \mathrm{ATENT} = \mu + \overline{\mathbf{x}}_{t=0}\boldsymbol{\delta} \end{cases} \tag{9}$$

Given these results, we can define the dose-response function of our model by simply averaging ATE(**x**, *t*) over **x**, that is:

Definition 2. *Formula of the dose-response function*. Given previous definitions and assumptions, we define the dose-response function associated to the treatment *t* as:

$$\mathrm{ATE}(t,w) = \mathrm{E}_{\mathbf{x}}\{\mathrm{ATE}(\mathbf{x},t,w)\} = w\cdot[\mathrm{ATET} + (h(t) - \overline{h}_{t>0})] + (1-w)\cdot\mathrm{ATENT} \tag{10}$$

or equivalently:

$$\mathrm{ATE}(t) = \begin{cases} \mathrm{ATET} + (h(t) - \overline{h}_{t>0}) & \text{if} \quad t > 0 \\ \mathrm{ATENT} & \text{if} \quad t = 0 \end{cases} \tag{11}$$

that is a function of the treatment intensity *t*. The estimation of equation (11) under different identification assumptions is the main purpose of next sections.

3. The regression approach

In this section we consider the conditions for a consistent estimation of the causal parameters defined in (3) and (4) and thus of the dose-response function in (11). What it is firstly needed, however, is a consistent estimation of the parameters of the potential outcomes in (1) – we call here "basic" parameters – as both ATEs and the dose-response function are functions of these parameters. In this direction, it is possible to state this proposition:

Proposition 2. *Baseline random-coefficient regression*. Under previous definitions and assumptions, and in particular the form of the potential outcomes in model (1), to be substituted into Rubin's potential outcome equation $y_i = y_{0i} + w_i(y_{1i} - y_{0i})$, the following random-coefficient regression model can be obtained (Wooldridge, 1997):

$$y_i = \mu_0 + w_i \cdot \text{ATE} + \mathbf{x}_i\boldsymbol{\delta}_0 + w_i \cdot (\mathbf{x}_i - \overline{\mathbf{x}})\boldsymbol{\delta} + w_i \cdot (h(t_i) - \overline{h}) + \eta_i \qquad (12)$$

where $\eta_i = e_{0i} + w_i \cdot (e_{1i} - e_{0i})$. Appendix A provides the proof. See A2.

The equation sets out in (12), provides the baseline regression for estimating the basic parameters (μ_0, μ_1, $\boldsymbol{\delta}_0$, $\boldsymbol{\delta}_1$, ATE) and then all the remaining ATEs. Both a semi-parametric or a parametric approach can be employed as soon as a parametric or a non-parametric form of the function $h(t)$ is assumed. In both cases, however, in order to get a consistent estimation of basic parameters, we need some additional hypotheses. We start

by assuming first *Unconfoundedness* or *Conditional Mean Independence (CMI)*, showing that it is sufficient to provide parameters' consistent estimation. Then we remove this assumption and introduce other identifying assumptions.

3.1 Estimation under unconfoundedness

Put very concisely, unconfoundedness states that conditional on the knowledge of the true exogenous confounders **x**, the condition for randomization is restored, and causal parameters become identifiable. In this subsection we set out this assumption in this form:

Assumption 2. *Unconfoundedness* (or CMI). Given the set of random variables $\{y_{0i}, y_{1i}, w_i, \mathbf{x}_i\}$ as defined above, the following equalities hold:

$$\mathrm{E}(y_{ji} \mid w_i, t_i, \mathbf{x}_i) = \mathrm{E}(y_{ji} \mid \mathbf{x}_i) \quad \text{with } j = \{0,1\}$$

CMI is a sufficient condition for identifying ATEs and the dose-response function in this context. Indeed, this assumption entails that, given the observable variables collected in **x**, both *w* and *t* are exogenous in equation (12), so that we can write the regression line of the response *y* simply as:

$$\mathrm{E}(y_i \mid w_i, t_i, \mathbf{x}_i) = \mu_0 + w_i \cdot \mathrm{ATE} + \mathbf{x}_i\boldsymbol{\delta}_\mathbf{0} + w_i \cdot (\mathbf{x}_i - \overline{\mathbf{x}})\boldsymbol{\delta} + w_i \cdot (h(t_i) - \overline{h}) \tag{13}$$

and Ordinary Least Squares (OLS) can be used to retrieve consistent estimation of all parameters:

Proposition 3. *Ordinary Least Squares (OLS) consistency*. Under assumption 1 and 2 (CMI), the error tem of regression (12) has zero mean conditional on (w_i, $\mathbf{x}_i$, t_i), i.e.:

$$\mathrm{E}\left(\eta_i \left| w_i, t_i, \mathbf{x}_i \right.\right) = \mathrm{E}\left(e_{0i} + w_i \cdot (e_{1i} - e_{0i}) \left| w_i, t_i, \mathbf{x}_i \right.\right) = 0 \tag{14}$$

thus implying that Eq. (14) is a regression model whose parameters can be consistently estimated by OLS. The proof is in Appendix A. See A3.

Once a consistent estimation of the parameters in (13) is obtained, we can estimate ATE directly from this regression, and ATET, ATENT and the dose-response function by plugging the estimated basic parameters into formula (9) and (10). This is possible because these parameters are functions of consistent estimates, and thus consistent themselves. Observe that standard errors for ATET and ATENT can be correctly obtained via bootstrapping (see Wooldridge, 2010, pp. 911-919).

In order to complete the identification of ATEs and of the dose-response function, we finally assume a parametric form for $h(t)$:

Assumption 3. Given the model in (1), assume a three-degree polynomial form for the function $h(t_i)$, i.e.:

$$h(t_i) = at_i + bt_i^2 + ct_i^3 \tag{15}$$

where a, b, and c are parameters to be estimated in regression (13).

Given previous assumption, equation (13) becomes:

$$\mathrm{E}(y_i \mid \mathbf{x}_i, w_i, t_i) = \mu_0 + \mathbf{x}_i\boldsymbol{\delta}_\mathbf{0} + w_i\mathrm{ATE} + w_i[\mathbf{x}_i - \overline{\mathbf{x}}]\boldsymbol{\delta} + w_i[(at_i + bt_i^{\,2} + ct_i^{\,3}) - (a\mathrm{E}(t_i) + b\mathrm{E}(t_i^{\,2}) + c\mathrm{E}(t_i^{\,3})]$$

that is:

$$\begin{aligned}\mathrm{E}(y_i \mid \mathbf{x}_i, w_i, t_i) = \mu_0 + \mathbf{x}_i\boldsymbol{\delta}_\mathbf{0} + w_i\mathrm{ATE} + w_i[\mathbf{x}_i - \overline{\mathbf{x}}]\boldsymbol{\delta} + a[t_i - \mathrm{E}(t_i)]w_i + b[t_i^{\,2} - \mathrm{E}(t_i^{\,2})]w_i + \\ + c[t_i^{\,3} - \mathrm{E}(t_i^{\,3})]w_i\end{aligned} \tag{16}$$

Under CMI, an OLS estimation of equation (16) produces consistent estimates of the parameters, we indicate as $\hat{\mu}_0, \hat{\boldsymbol{\delta}}_\mathbf{0}, \hat{\mathrm{ATE}}, \hat{\boldsymbol{\delta}}, \hat{a}, \hat{b}, \hat{c}$. With these parameters at hand, we can finally consistently estimate the dose-response function as:

$$\widehat{\mathrm{ATE}}(t_i) = w[\widehat{\mathrm{ATET}} + \hat{a}(t_i - \frac{1}{N}\sum_{i=1}^{N} t_i) + \hat{b}(t_i^2 - \frac{1}{N}\sum_{i=1}^{N} t_i^2) + \hat{c}(t_i^3 - \frac{1}{N}\sum_{i=1}^{N} t_i^3)] + (1-w)\widehat{\mathrm{ATENT}} \tag{17}$$

where:

$$\widehat{\mathrm{ATET}}(t_i) = \widehat{\mathrm{ATE}}(t_i)_{t_i>0}$$

A simple plot of the curve $\widehat{\mathrm{ATE}}(t_i)_{t_i>0}$ over the support of *t* returns the pattern of the dose-response function. Moreover, for each level of the dose *t*, it is also possible to calculate the α-confidence interval around the dose-response curve.

Proposition 4. *Analytical standard error for the dose-response function*. 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 dose-response function is equal to[4]:

$$\hat{\sigma}_{\hat{\mathrm{ATE}}(t)} = \left\{ T_1^2\hat{\sigma}_a^2 + T_2^2\hat{\sigma}_b^2 + T_3^2\hat{\sigma}_c^2 + 2T_1T_2\hat{\sigma}_{a,b} + 2T_1T_3\hat{\sigma}_{a,c} + 2T_2T_3\hat{\sigma}_{b,c} \right\}^{1/2} \tag{18}$$

Proof in Appendix A. See A4.

This implies that the α-confidence interval of $\hat{\mathrm{ATE}}(t)$ for each t is given by:

$$\left\{ \hat{\mathrm{ATE}}(t) \pm Z_{\alpha/2} \cdot \hat{\sigma}_{\hat{\mathrm{ATE}}(t)} \right\}$$

that can be usefully plotted along the dose-response curve for visually detecting the statistical significance of the treatment effect along the support of the dose t.

3.2 Estimation under treatment endogeneity

When w (and thus t) are endogenous the CMI assumption no longer holds, and the OLS estimate of regression (16) becomes biased. This occurs because the orthogonality condition implied by equation (14) fails, so that:

$$\mathrm{E}\left(\eta_i \middle| w_i, t_i, \mathbf{x}_i\right) = \mathrm{E}\left(e_{0i} + w_i \cdot (e_{1i} - e_{0i}) \middle| w_i, t_i, \mathbf{x}_i\right) \neq 0 \tag{19}$$

[4] 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.

where it is clear that inequality depends on the endogeneity of w_i (and t_i), being $\mathbf{x}_i$ assumed to be pre-determined. In such a case, however, an Instrumental-Variables (IV) estimation may be implemented to restore consistency, provided that $e_1=e_0$ (Wooldridge, 2010, pp. 942-943)[5]. Such condition excludes the presence of unobservable heterogeneity, while preserving observable heterogeneity. In order to implement IV in such a setting, we first need to express the previous model in a semi-structural form, that is:

Assumption 4. *Semi-structural form of model (1)*. Given Assumption 1, and the potential outcome model, we can write that[6]:

[5] Assuming $e_1=e_0$ leads to IV consistency without the need to introduce further distributional assumptions on the error terms as in the traditional Heckit model. Given instruments **z**, an alternative IV identification assumption could be $E(w(e_1 - e_0) \mid \mathbf{x}, \mathbf{z}) = E(e_1 - e_0 \mid \mathbf{x}, \mathbf{z})$ which is, however, strongly ad hoc (Wooldridge, 2010, pp. 944).

[6] It is worth stressing that our model assumes that the (public) agency stepwise decides either: (i) who is entitled to receive the treatment, and (ii) the level of the treatment". It means that this model does not account for the dose to be agents' choice. Furthermore, it is assumed that the level of the received treatment and the level of the dose actually employed by units are the same. When the level of treatment actually exploited is however lower than the treatment actually received, thereby implying that dose becomes a strategic variable for units, then the observed dose becomes "endogenous" (as "measured with error"). In this case a proper IV procedure can restore consistent treatment effects' estimation, despite the presence of this sort of "imperfect compliance" with treatment. For an interesting discussion on this issue, see the paper by Angrist (2006).

.

$$
\begin{aligned}
&y_i = \mu_0 + \mathbf{x}_i\boldsymbol{\delta_0} + w_i\text{ATE} + w_i[\mathbf{x}_i - \overline{\mathbf{x}}]\boldsymbol{\delta} + aw_iT_{1i} + bw_iT_{2i} + cw_iT_{3i} + \eta_i \\
&w_i = \begin{cases} 1 & \text{if} \quad w_i^* > 0 \\ 0 & \text{if} \quad w_i^* \le 0 \end{cases} \\
&t_i = \begin{cases} t_i' & \text{if} \quad w_i^* > 0 \\ - & \text{if} \quad w_i^* \le 0 \end{cases}
\end{aligned} \tag{20}
$$

where: T_{1i}=t_i-E(t_i), T_{2i}=t_i^2-E(t_i^2) and T_{3i} = t_i^3-E(t_i^3); w_i^* represent the latent unobservable counterpart of the binary variable w_i (for instance, w_i^* might be seen as the net benefit - cost minus return - of an agency choosing to finance specific subjects); t_i is fully observed only when w_i=1 (and t_i=t_i'), otherwise it is supposed to be unobserved (although put equal to zero). By making explicit in linear form the reduced-form equations for w_i^* and t', the previous model may be re-written as follows:

$$
\begin{cases}
y_i = \mu_0 + \mathbf{x}_i\boldsymbol{\delta_0} + w_i\text{ATE} + w_i[\mathbf{x}_i - \overline{\mathbf{x}}]\boldsymbol{\delta} + w_iT_{1i} + bw_iT_{2i} + cw_iT_{3i} + \eta_i & (21.1) \\
w_i^* = \mathbf{x}_{w,i}\boldsymbol{\beta}_w + \varepsilon_{wi} & (21.2) \\
t_i' = \mathbf{x}_{t,i}\boldsymbol{\beta}_t + \varepsilon_{ti} & (21.3)
\end{cases}
$$

where: $\mathbf{x}_{w,i}$ and $\mathbf{x}_{t,i}$ are two sets of exogenous regressors, and ε_{wi}, ε_{ti} and η_i are error terms supposed to be freely correlated one another with zero unconditional means. Equation (21.2) – the *selection* equation – defines the regression explaining the net benefit indicator w^*. The vector of covariates $\mathbf{x}_{w,i}$ are the selection criteria used, for instance, by an agency to set the treated and untreated group. In turn, equation (21.3) – 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 $\mathbf{x}_{t,i}$ are those exogenous variables thought of as determining the treatment level.

In equation (21.1), w_i and T_{1i}, T_{2i} and T_{3i} are endogenous, being these latter ones functions of the endogenous t. This entails that the problem in this system is with parameters' identification. In general, with two endogenous variables, the identification of linear systems of equations would require the availability of at least two instrumental variables, one for w and one for t (i.e., just-identified setting).

Assumption 5. *Identification of system (21)*. Given system (21), we assume to know two exogenous variable $z_{w,i}$ and $z_{t,i}$ supposed to be: (*i*) correlated with w_i^* and t_i', respectively; (*ii*) uncorrelated with ε_{wi}, ε_{ti} and η_i. As such, $z_{w,i}$ and $z_{t,i}$ behave as instrumental variables.

Assumption 5 leads naturally to the following specification of the exogenous confounders in system (21):

$$\begin{aligned} \mathbf{x}_{w,i} &= [\mathbf{x}_i; z_{w,i}] \\ \mathbf{x}_{t,i} &= [\mathbf{x}_i; z_{t,i}] \end{aligned} \tag{22}$$

so that a full specified model - with all the equations depending on the same exogenous $\mathbf{x}_i$ - is considered, where $z_{w,i}$ and $z_{t,i}$ are the instrumental variables directly correlated with the selection and the level of treatment, but directly uncorrelated with the level of the outcome.

Nevertheless, since equation (21.2) is a latent regression model, ultimately estimated through a logit or probit regression, parameters of system (21) may be identified just by means of one instrumental variable, and in particular by the use of $z_{t,i}$. This depends on the non-linearity of the logit or probit regression, producing orthogonal projections (i.e., predictions) that are not perfectly collinear with $\mathbf{x}_i$ in equation (21.1). However, relying on two (instead of one) instruments increases the precision of

parameters' estimation in (21.1), as the degree of collinearity is reduced and the variance of IV estimators shrinks accordingly.

Practical estimation of system (21) starts from recognizing that the two last equations of system (21) – i.e., (21.2)-(21.3) – represents a bivariate sample-selection model or type-2 tobit model (Heckman, 1979). Generally, such a model is estimated by invoking some distributive assumptions regarding the error terms.

Assumption 6. We assume that the error terms in (21.2) and (21.3) are jointly normally distributed and homoskedastic:

$$\begin{bmatrix} \varepsilon_{wi} \\ \varepsilon_{ti} \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}; \begin{bmatrix} 1 & \sigma_{wt} \\ \sigma_{wt} & \sigma_t^2 \end{bmatrix}\right]$$

where the normalization σ_w=1 is used because only the sign of w_i^* is observed.

Given this additional assumption, all the ingredients to provide a procedure for estimating system (21) consistently are available.

Proposition 5. *Consistent estimation of system (21)*. Under assumption 4, 5 and 6, and by assuming (22) to hold, the following procedure provides a consistent estimation of the parameters of system (21), and in particular of equation (21.1):

1. *First*: estimate equations (10.2)-(10.3) jointly by a type-2 tobit model.

Comment. As said, this can be achieved by a Heckman two-step procedure (Heckman, 1979). The Heckman two-step procedure performs: a probit of w_i on $\mathbf{x}_{w,i}$ in the first step using only the N_1 selected observations; and an OLS regression of t_i' on $\mathbf{x}_{t,i}$, augmented by the Mills' ratio obtained from the probit in the second step, using all the N observations since predictions are made also for the censored data. However, because of the errors' joint normality, a maximum-likelihood (ML) estimation can be also employed; ML leads to more efficient estimates of $\boldsymbol{\beta}_w$ and $\boldsymbol{\beta}_t$.

2. *Second*: compute the predicted values of w_i (i.e. $\hat{p}_{wi}$) and t_i (i.e. $\hat{t}_i$) from the previous type-2 tobit estimation, and then perform a two-stage least squares (2SLS) for equation (21.1) using as instruments the following exogenous variables ($\mathbf{x}_i, \hat{p}_{wi}, \hat{p}_{wi}[\mathbf{x}_i - \overline{\mathbf{x}}], \hat{p}_{wi}\hat{T}_{1i}, \hat{p}_{wi}\hat{T}_{2i}, \hat{p}_{wi}\hat{T}_{3i}$).

Comment. This 2SLS approach provides consistent estimation of the basic coefficients μ_0, $\boldsymbol{\delta}_0$, ATE, $\boldsymbol{\delta}$, a, b, c (Wooldridge, 2010, pp. 937-951)[7].

3. *Third*: once previous procedure estimates consistently the basic parameters in system (21), the causal parameters of interest - ATEs and the dose-response function - can be consistently estimated by the same plug-in approach used for the OLS case.

Proof in Appendix A. See A5.

[7] Observe that instruments used in the 2SLS are based on the orthogonal projection of w_i and t_i on the vector space generated by all the exogenous variables of system (21).

The question of identification of parameters in the system of equations (21) using the previous procedure is a bit trickier than it may appear at first glance. As suggested by Wooldridge (2002, p. 613), one can be selective in establishing which subset of variables contained in the vector **x** of equation (21.1) has to interact with the treatment w in the fourth term of the RHS of (21.1). Suppose to call such subset of the **x**-variables as $\mathbf{x}_{\text{hetero}}$. In the discussion on identification presented so far, we assumed that:

$$dim(\mathbf{x}_{\text{hetero}}) = dim(\mathbf{x})$$

In this case, identification requires – as argued – at least one instrument, and in particular the instrument $z_{t,i}$. As said, this depends on the nonlinearity of the probit/logit model, so that the instrument $z_{w,i}$ turns out to be unnecessary for identification – although its use could increase estimation efficiency. This follows exactly the same argument one typically invokes in the standard Heckman selection model (Heckman, (1979), where instruments are not necessary for identifying ATEs because of the first-step probit estimation. As a conclusion, previous procedure can identify all the causal parameters (including the dose-response function) by means of just one instrument, $z_{t,i}$.

Nonetheless, assume instead to have:

$$dim(\mathbf{x}_{\text{hetero}}) < dim(\mathbf{x})$$

i.e., the set of variables included into $\mathbf{x}_{\text{hetero}}$ contains just "some" of the variables in **x** (not all). In this case, by construction, one does not need any instrument to identify the proposed model. This depends on the fact that, by restricting the number of variables

supposed to interact with *w* (i.e., having heterogeneous response) in the two states (treated vs. untreated), we introduce "exclusion restrictions" which reduce the number of endogenous variables (as the interactions are in turn endogenous, being them function of *w* which is supposed to be endogenous by assumption), thus "helping" to reach identification.

This means that, by just excluding one of the variables in **x** from the set $\mathbf{x}_{\text{hetero}}$, we can identify the model with no instruments. This is appealing, but has some costs in terms of reduced estimates' efficiency. However, since some of the variables **x** are probably not subject to observable heterogeneity, the IV model can identify ATEs without using instruments, as the standard Heckman selection model does. Of course, part of the identification is based on the type-2 tobit model assumed in the first step, which relies on the bivariate normality of the errors. We will come back to this issue in section 4.2[8].

[8] It is not clear at this stage which is the link between the definition of the Local Average Treatment Effect (LATE) as proposed by Imbens and Angrist (1994), and the IV approach as proposed in this paper. The problem is that LATE identifies the casual effect of *w* on *y* in a setting where the instrument *z* is binary. Although extension to the case in which *w* can be multi-valued and more than one binary *z* is available have been provided (Angrist and Pischke, 2008, 173-186), no comparable findings are found so far in the literature for the case in which both the treatment and the instrument take values on a continuous support. The use of instrumental variables under heterogeneous effects will however identify a causal effect for a subpopulation, and not for the (complete) population. This is irrespective of the precise characterization of the subpopulation (e.g., compliers in the binary setting). This certainly represents a trade-off when moving to the use of IVs, when there is effect heterogeneity based on unobservables.

3.3 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:

$$\mathrm{ATE}(t,\Delta) = \mathrm{E}[y(t+\Delta) - y(t)] \tag{23}$$

Equation (23) identifies the average treatment effect between two states (or levels of treatment): t and $t+\Delta$. Given a level of $\Delta = \overline{\Delta}$, we can get a particular $\mathrm{ATE}(t,\overline{\Delta})$ that can be seen as the "treatment function at $\overline{\Delta}$". Observe that the standard ATE(t) is obtained from (23) when t=0.

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

$$\begin{aligned}&\mathrm{E}(y \mid t) = \mathrm{E}_{\mathbf{x},w=1}\{\mathrm{E}(y \mid \mathbf{x}, w, t)\} = \mathrm{E}_{\mathbf{x},w=1}\{\mu_0 + \mathbf{x}\boldsymbol{\delta_0} + w\mathrm{ATE} + w[\mathbf{x} - \overline{\mathbf{x}}]\boldsymbol{\delta} + w[h(t) - \overline{h}]\} = \\ &\mu_0 + \overline{\mathbf{x}}\boldsymbol{\delta_0} + \mathrm{ATE} + (h(t) - \overline{h})\end{aligned} \tag{24}$$

Therefore:

$$\begin{aligned}&\mathrm{E}(y \mid t+\Delta) - \mathrm{E}(y \mid t) = [\mu_0 + \overline{\mathbf{x}}\boldsymbol{\delta_0} + \mathrm{ATE} + (h(t+\Delta) - \overline{h})] - [\mu_0 + \overline{\mathbf{x}}\boldsymbol{\delta_0} + \mathrm{ATE} + (h(t) - \overline{h})] = \\ &h(t+\Delta) - h(t)\end{aligned} \tag{25}$$

that is:

$$\mathrm{ATE}(t,\Delta) = \mathrm{E}[y(t+\Delta) - y(t)] = h(t+\Delta) - h(t) = [\, a(t+\Delta) + b(t+\Delta)^2 + c(t+\Delta)^3\,] - [\, at + bt^2 + ct^3\,] \tag{26}$$

and an estimation is thus given by:

$$\widehat{\mathrm{ATE}}(t,\Delta) = \hat{a}(t+\Delta) + \hat{b}(t+\Delta)^2 + \hat{c}(t+\Delta)^3 - [\hat{a}t + \hat{b}t^2 + \hat{c}t^3] \tag{27}$$

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

4. Applications[9]

4.1 Application 1: the impact of public support on business R&D

In this section, we present an application of our model on real data. We aim at estimating the effect of public research and development (R&D) support on company R&D performance. The level of the public support to R&D, in fact, is a typical continuous treatment variable showing a large number of zeros for non supported companies and a positive value for all supported units. As dataset, we employ the 8th, 9th and 10th Unicredit surveys collecting a large body of information on various characteristics and activities of a sample of Italian companies, including innovation and R&D (public and private) financing for around 5,000 companies in each wave. The timing is: 1998-2000 for the 8th survey, 2001-2003 for the 9th survey, and 2004-2006 for

[9] A Stata implementation of the model presented in previous sections can be found in Cerulli (2015).

the 10th one. All surveys are built by stratifying on sector, size and location, thus being representative of Italian manufacturing companies with more than 10 employees.

The three surveys are then combined in a unique repeated cross-section of 14,106 companies, since building a longitudinal dataset (or panel) would have caused a sharp reduction in the sample size (only 451 companies appear in all the three surveys). Furthermore, since only relatively few businesses present information on R&D financing, being this section of the questionnaires very rich of missing values, exploiting a repeated cross-section guarantees a larger (final) sample size. The final dataset is then merged with companies' balance sheet data coming from the AIDA archive[10].

Table 1 reports the model's specification for this application. Here a description of the outcomes, binary treatment, treatment level (or dose), and control covariates is compactly reported. This specification of the outcome equation comes from a widely accepted view of the main drivers of company R&D performance as maintained in the R&D policy evaluation literature (David et al. 2000; Cerulli and Potì, 2012).

In each survey, variables are calculated through a three-year average so to have common time consistency. Therefore, given the time structure of the Unicredit surveys, we perform a three-period analysis with each period made of an average over three years. In sum, in this application we cover nine years, from 1998 to 2006.

[10] AIDA is a commercial database on Italian firms maintained by Bureau van Dijk. It gives balance-sheet, income statement and other information, such as location, sector, year of incorporation, ownership and equity participations in other firms, covering a 10-year time window. More information on this dataset are available at this URL: http://www.bvdinfo.com/it-it/our-products/company-information/national-products/aida.

An usual problem in the impact evaluation of company R&D public support is the difficulty in singling out good candidates as instrumental variables. Our application shares this limitation with the previous literature. Therefore, in this exercise we only apply the OLS control-function approach, by leaving the IV approach to be used in the application presented in the next section.

Before commenting the results, it seems firstly useful to inspect into the dataset we obtain after combining and merging these different data sources. Our regression's specification, unfortunately, leads to a huge drop of observations. Indeed, due to a great number of crossing missing values and after deleting influential observations, we are left with a sample of 932 observations. Nevertheless, Table 2 shows that the representativeness of the population is quite well kept in the sample both in terms of sector and location. As for the size, on the contrary, we have a higher presence of companies having between 51-250 employees and a lower presence of smaller firms. This might generate a little bias towards less financial constrained companies, thus probably driving down the average effect of the R&D support considered here.

Another important descriptive statistics is the distribution of the R&D publicly supported, measured as the share of total firm R&D expenditure. Table 3 shows that in our sample this average share is around 40% with a median equal to 30%. It means that, among the supported companies, a large quota of R&D is generally covered by public financing.

<TABLE 1 - HERE>

<TABLE 2 - HERE>

The number of supported companies is rather small, 235 in the final sample: this is due to the fact that very few companies respond to the financing section of the Unicredit questionnaires. Nevertheless, at least in Italy, Unicredit surveys are the only freely available datasets providing the level of the R&D public support at firm level (Cerulli and Potì, 2012).

<TABLE 3 - HERE>

Given this picture, we can assess the impact of public R&D support on the four outcome measures reported in Table 1. We have to observe, however, that the most relevant outcome for our purposes is the level of "net R&D expenditure" (R&D expenditure *minus* the subsidy received) as it returns the actual amount of *additional* R&D that a company has been able to perform. Nevertheless, also the effect on "gross" (or "total") R&D outlay will be estimated. Moreover, as scale effects can be relevant even if controlling for firm size as we do here, we calculate also the effect on both gross and net R&D expenditure either on turnover and per employee. Results are set out in Table 4 and 5, for all variables, although it seems more interesting the graphical pattern of the estimated dose-response function and of the distributions of ATE($\mathbf{x}$,t), ATET($\mathbf{x}$,t) and ATENT($\mathbf{x}$,t). These are visible in Figure 1.

<TABLE 4 - HERE>

<TABLE 5 - HERE>

Results in Table 4 show that public support to both gross and net R&D is effective at least at the 5% significant level. Quite surprisingly, the effect on net R&D is a bit higher, around 325 thousand euro. Other significant predictors are s*ize* (measured as number of employees, with a positive sign), *cash-flow* (although with a negative sign) and belonging to a group of firms (with a positive value). In terms of coefficients' magnitude, measured by standardized coefficients, Table 5 reveals that *size* is the most relevant predictor followed by the *Treatment* having however a coefficient three times lower.

In the R&D literature, it is well recognized that company size has generally a great impact on R&D spending; thus, having found a positive and significant effect of R&D support even by controlling for company size, suggests that public financing has been decisively effective in fostering the level of both net and gross R&D expenditure. Nevertheless, when considering both gross and net R&D per capita and R&D intensity (obtained through dividing by firm turnover) results show no significant (although positive) effect of R&D public support. This might indicate that the additional level of R&D induced by the public support has been not comparatively higher than the growth in company employees and turnover. As such, previous conclusions on support's effectiveness might be carefully reconsidered, as scale-neutral R&D performance indicators do not show such a result. Anyway, if the level of R&D is the main policy objective, moderately optimistic conclusions on the achievements of the policy thus evaluated can be drawn.

Figure 1 shows the kernel estimation of the distribution of ATE($\mathbf{x}$,t), ATET($\mathbf{x}$,t) and ATENT($\mathbf{x}$,t), and the plot of the dose-response function with 95% confidence intervals. As for the distributions, it is immediate to see that the net R&D performance

shows, in each graph, a much more disperse distribution for ATET(**x**,*t*) compared with ATE(**x**,*t*) and ATENT(**x**,*t*). Moreover, ATET(**x**,*t*) appears much more concentrated on lower values, thus indicating that the effect on treated units seems surprisingly not only less regular, but also weaker for treated than for untreated units. This might question the selection process adopted by the public agency (although, on average, differences are not strong).

More interesting for the aim of this paper is the pattern of the dose-response functions. As for both gross and net R&D, it is easy to see that the dose-response function has a negative slope with significant confidence intervals lying in between (0;20] for gross R&D spending, and in between (0;15] for net R&D. This result is striking, as it says that the overall positive effect of the policy found in the previous regression results (Table 4 and 5) is mainly driven by those supported companies getting a comparatively lower share of R&D covered by public support (no more than around 20% or 15% for gross and net R&D respectively). A similar finding has been found in Marino and Parrotta (2010), using the Hirano and Imbens' approach on company R&D support in Denmark. No effect seems to emerge for higher shares of publicly financed R&D, as the dose-response functions decrease slowly with very large confidence intervals. Interestingly, company net R&D expenditure becomes negative around a threshold of 40%: this finding might have remarkable policy implications.

Results on R&D per capita and R&D intensity seem similar and in tune with previous regression output: confidence intervals are almost uniformly large, with the pattern of net R&D appearing decisively more decreasing than that of gross R&D.

<FIGURE 1 - HERE>

4.2 Application 2: the impact of job tenure on wages

In this section we provide an illustrative example on how applying to real data the IV approach developed in section 3.2. We consider the dataset "nlswork" coming from the US National Longitudinal Survey (NLS) reporting job and personal information on young women aged between 14 and 26, based on a series of interviews carried out from 1968 to 1988[11]. This dataset contains information on women's labor conditions such as wages, educational level, race, marital status, etc.. We are interested in estimating the impact of the variable "tenure" (job tenure) on "wage" (measured as the logarithm of wages in dollars per hour) conditional on a series of covariates (i.e., observable confounders). The variable "tenure" is a good candidate to be exploited as continuous-treatment (i.e., dose) as it presents a spike at zero with 1,248 out of 28,534 observations (a share of around 4,5 %) reporting a level of job tenure equal to zero (with 433 observations presenting missing value). The description of the dataset is illustrated in Table 6.

In regressing wage on tenure (and controls), a possible estimation problem might arise if one assumes that unobservable shocks affecting a woman's wage also affect her tenure, thus making job tenure endogenous. In such a case, relying on an OLS control-function regression – as we did in the application of section 4.1 – might lead to biased results. In order to recover consistent estimates, we can however exploit the IV procedure presented in section 3.2.

This application is based on that illustrated in StataCorp (2013, p. 936) where a specific specification of the model is used to fit a standard IV regression. This application assumes as control variables: woman's age and age squared, birth year, and

[11] The dataset can be freely downloadable at this URL: http://www.stata-press.com/data/r13/nlswork.dta.

level of education. As for the instruments, woman's union status ("union"), and a dummy indicating whether she is married with a spouse present at home ("msp") are used. Indeed, the event to present a union membership, and that of living with a spouse present at home are thought to be reasonably correlated with job tenure, while should not have a direct effect on wages. In the StataCorp's example, the IV estimate of ATE results in a magnitude which is 2.83 times that of the OLS. We take such a value as a heuristic benchmark to assess the reliability of our IV approach.

<TABLE 6 – ABOUT HERE>

Before fitting this model and thus drawing the dose-response function, however, we first generate the binary treatment variable we call "treatment", and the continuous-treatment (i.e., the dose) ranging from 0 and 100 we call here "tenure2". This way, we have all the ingredients to apply our IV procedure. We start with estimating the model by OLS, and then by IV when we assume treatment endogeneity.

According to our results, OLS estimates return an ATE equal to 0.09 which is highly significant with a standard error equal to 0.013, t-test of 6.97, sample size of 28,099, and R-squared of 0.30. Our IV approach returns instead a non-significant estimate of ATE of around 2.38 which is around 26 times larger than the OLS estimate. It is an unreliable result which deserves further inspection. We can in fact show that such overestimation and very low precision of our IV estimate of ATE can be explained by a rather strong non-normality of the "tenure2" variable, whose empirical distribution is reported in panel (a) of Figure 2. As clearly visible, the distribution of "tenure2"

appears to be closer to a Power Law (as, for instance, a Pareto) distribution, which is highly skewed and thus well far from having a Normal shape.

<FIGURE 2 – ABOUT HERE>

In order to fix such problem, a possible solution might be that of using a logarithmic transformation of "tenure2" which should reduce skewness. We call the variable thus obtained "tenure3", whose distribution is reported in panel (b) of Figure 2. By excluding the evident high frequency (or "spike") at zero, the rest of this distribution is still far from being Normal, and the IV estimation of ATE thus obtained results in a very imprecise value which is around 33 times that of the (significant) OLS estimation. Hence, even in this case, our IV procedure returns unreliable estimates of the average treatment effect. As said, we suspect it depends on a weak compliance with the normality assumption of the logarithm of job tenure. Therefore, in order to see what happens if we force our data to more tightly comply with the Normality assumption, we truncate the distribution of "tenure3" by deleting all the observations between zero (not included) and a dose level of 42. This way, we eliminate its non-Normal part. This truncation results in a new distribution of the log of job tenure visible in panel (c) of Figure 2. This distribution is evidently closer to a Normal one except, of course, for the truncation on its left tail. Although a perfect compliance with the Normal distribution is not fully achieved also in this case, IV estimates become now 12 times larger than OLS estimates, thus considerably improving the reliability of our IV procedure. This finding is sufficient to prove that, unlike our OLS approach, the IV procedure proposed in

section 3.2 seems to be quite sensitive to departure from Normality. A series of Monte Carlo experiments conducted on this model confirm such finding.

Taking the distribution of the truncated "tenure3" as the most reliable variable, Tables 7 and 8 show, respectively, the results from the first and the second step of our IV procedure. The still large imprecision of the IV estimator of the treatment effect (with magnitude 1.66, and standard error 3.19), might be however due not only to the weak Normality of the truncated "tenure3", but also to the weak nature of the instrument "mps" which seems poorly related - in a multivariate sense – to the endogenous treatment.

<TABLE 7 – ABOUT HERE>

<TABLE 8 – ABOUT HERE>

Although our IV point estimation is imprecisely estimated, the shape of the dose-response function seems reasonably reliable. Figure 3 illustrates the pattern of the dose-response function of "job tenure" on "wage" when using OLS (panel (a)) and IV (panel (b)) with the truncated "tenure3" variable. Both curves clearly show the presence of the truncation, and confirm the growing pattern of ATE(t) – i.e., the dose-response curve – as a function of the dose (i.e., tenure). As expected, it is also evident the larger confidence intervals of IV compared to OLS estimates, which confirms what found above for the IV point estimate of ATE.

<FIGURE 3 – ABOUT HERE>

By assuming treatment endogeneity, we can thus conclude that the "true" dose-response function of "wages" on "tenure" is probably located somewhere above the OLS and below the IV dose-response functions. It is also clear that such relation is an increasing one, a fact confirmed also by our IV approach which returns - at least in this case - a robust result.

5. Conclusion

This paper has presented an original econometric model for estimating a dose-response function through a regression approach where: (i) treatment is continuous with a spike at zero, (ii) individuals may react heterogeneously to observable confounders, and (iii) selection-into-treatment may be endogenous. This model tries to overcome some limitations of previous counterfactual models with continuous treatment, and in particular the one proposed by Hirano and Imbens (2004), by taking into account also the presence of untreated units (i.e., by modeling zero-treatment), and the presence of potential treatment endogeneity.

Two estimation procedures have been proposed for this model: one based on OLS under Conditional Mean Independence (or CMI), and one based on Instrumental-variables (IV) assuming selection endogeneity.

Two applications to real data have been set out. In the first application, we assessed the effect of public R&D support (measured as a share of total company R&D expenditure) on business gross and net R&D outlay using the proposed OLS procedure. Results seem to shed new light on the relation between public support and its effect on

company R&D behavior[12]. In particular, the proposed model allows to look at the pattern of the policy effect over treatment intensity, thus going beyond the typical average-effect analysis. Recovering the pattern of the dose-response function, along with the plot of its confidence intervals, can be used for a better inspection into the causal relation between the policy instrument and the policy target. This is a relative advantage of using dose-response models than traditional evaluation approaches.

The second application has explored the relation between job tenure and wages at individual level. For the sake of comparison, by drawing on an existing example applying a traditional IV approach, this application confirms an increasing relation between job tenure and wages, and shows that, generally speaking, the normality assumption of the first-step regression of our IV approach is relevant for results to be stable and precise. Providing an IV procedure in the case of an endogenous dose-response function able to relax the Normality assumption of the fist-step seems thus a challenging agenda for next research.

[12] In relation to this, see also Cerulli (2010) for an econometric review of the treatment models used in this field of study.

References

Abadie, A., and Imbens, G. W. (2012).Matching on the estimated propensity score. Harvard University and National Bureau of Economic Research. http://www.hks.harvard.edu/fs/aabadie/pscore.pdf.

Angrist J., Pischke S. (2008), *Mostly Harmless Econometrics*, Princeton University Press.

Angrist, J.D. (2006), Instrumental variables methods in experimental criminological research: what, why and how, *Journal of Experimental Criminology*, 2, 1, 23-44.

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.A., Flores-Lagunes, A., and Mattei, A. (2014), A Stata package for the application of semiparametric estimators of dose-response functions, *The Stata Journal*, 14, 3, 580-604.

Cerulli, G. (2015), ctreatreg: Command for fitting dose-response models under exogenous and endogenous treatment, *The Stata Journal*, 15, 4, 1019-1045.

Cerulli, G. (2014), ivtreatreg: A command for fitting binary treatment models with heterogeneous response to treatment and unobservable selection, *The Stata Journal*, 14, 3, 453-480.

Cerulli, G. (2010), Modelling and measuring the effect of public subsidies on business R&D: a critical review of the econometric literature, *Economic Record*, 86, 421-449.

Cerulli, G., and Potì, B. (2012), Designing ex-post assessment of corporate RDI policies: conceptualization, indicators, and modeling, *World Review of Science, Technology and Sustainable Development*, 9, 2/3/4.

David, P.A., Hall, B.H., and Toole, A.A. (2000), Is public R&D a complement or substitute for private R&D? A review of the econometric evidence, *Research Policy*, 29, 497–529.

Dehejia, R., and Wahba, S. (2002). Propensity score–matching methods for nonexperimental causal studies. The Review of Economics and Statistics, 84, 151–161.

Guardabascio, B. and Ventura, M. (2014), Estimating the dose-response function through a generalized linear model approach, *The Stata Journal*, 14, 1, 141-158.

Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica: Journal of the 1768 Econometric Society, 47, 153–161.

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.

Imbens, G.W. and Angrist, A.D. (1994), Identification and Estimation of Local Average Treatment Effects, *Econometrica*, Vol. 62, No. 2, pp. 467-475.

Marino, M., and Parrotta, P. (2010), "Impacts of public funding to R&D: evidence from Denmark", Paper presented at the DRUID Summer Conference 2010 on "Opening Up Innovation: Strategy, Organization and Technology", Imperial College London Business School, June 16-18.

Robertson, C., Boyle, P., Hsieh, C.C., Macfarlane, G.J., and Maisonneuve, P. (1994), Some statistical considerations in the analysis of case-control studies when the exposure variables are continuous measurements, *Epidemiology*, 5,164-170.

Rosenbaum, P. R., and Rubin, D. B. (1984). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.

Royston, P., and Sauerbrei, W. (2008), *Multivariable Model-building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. Wiley: Chichester.

Royston, P., Sauerbrei, W., and Becher H. (2010), Modelling continuous exposures with a 'spike' at zero: a new procedure based on fractional polynomials. *Statistics in Medicine*, 29, 1219-27.

StataCorp (2013), Stata base reference manual - Release 13. College Station, TX: StataCorp LP.

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. (2002): *Econonometric analysis of cross section and panel data*, MIT Press.

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.

Appendix A.

A1. Proof of proposition 1. *Formula of ATE*.

Define z and w being two random variables. LIE states that:

$$\mathrm{E}(z) = \mathrm{E}_w\{\mathrm{E}_z\{z \mid w)\} \tag{A1.1}$$

Assume that:

$$z = (y_1 - y_0 \mid \mathbf{x}, t) \tag{A1.2}$$

and define:

$$\mathrm{ATE}^* = \mathrm{E}(y_1 - y_0 \mid \mathbf{x}, t) = \mathrm{E}(z) \tag{A1.3}$$

Now, define:

$$\mathrm{ATE}^*(w) = \mathrm{E}(z \mid w) \tag{A1.4}$$

Using these definitions we have that:

$$\mathrm{ATE}^* = \mathrm{E}(y_1 - y_0 \mid \mathbf{x}, t) = \mathrm{E}(z) = \mathrm{E}_w\{\mathrm{E}(z \mid w)\} = \mathrm{E}_w\ \{\mathrm{ATE}^*(w)\} \tag{A1.5}$$

Now, ATE*(w) is a *random variable* with this distribution:

$$\mathrm{ATE}(w)=\begin{cases}\mathrm{ATE}(w{=}1) = \mathrm{ATET} & pr(w{=}1)\\ \mathrm{ATE}(w{=}0) = \mathrm{ATENT} & 1\text{-}\ pr(w{=}1)\end{cases} \tag{A1.6}$$

implying that:

$$\begin{aligned}\text{ATE}^* = \text{E}_w\{\text{ATE}^*(w)\} = \text{ATE}^*(w{=}1)\, pr(w{=}1) + \text{ATE}^*(w{=}0)\,[1 - pr(w{=}0)] = \\ \text{E}(y_1 - y_0 \mid \mathbf{x},\, t{>}0)\, pr(w{=}1) + \text{E}(y_1 - y_0 \mid \mathbf{x},\, t{=}0)\,[1 - pr(w{=}0)] = \text{ATE}(\mathbf{x},t)\end{aligned} \tag{A1.7}$$

Finally, by applying LIE again, we get:

$$\text{ATE} = \text{E}(\text{ATE}^*) = \text{E}_{\mathbf{x},t}[\text{ATE}(\mathbf{x},t)] \tag{A1.8}$$

implying that:

$$\text{ATE}(\mathbf{x},t,w) = w[\mu + \mathbf{x}\boldsymbol{\delta} + h(t)] + (1-w)[\mu + \mathbf{x}\boldsymbol{\delta}] \tag{A1.9}$$

□

A2. Proof of proposition 2. *Baseline random-coefficient regression*.

Substitute the potential outcomes' form of model (1) into Rubin's potential outcome equation $y = y_0 + w(y_1 - y_0)$, thus getting:

$$y = y_0 + w(y_1 - y_0) = (\mu_0 + g_0(\mathbf{x}) + e_0) - w[(\mu_1 + g_1(\mathbf{x}) + h(t) + e_1) - (\mu_0 + g_0(\mathbf{x}) + e_0)] \tag{A2.1}$$

Then, by collecting the various arguments of this equation, 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 that $g_1(\mathbf{x}) = \mathbf{x}\boldsymbol{\delta}_1$ and $g_0(\mathbf{x}) = \mathbf{x}\boldsymbol{\delta}_0$ as done in the text, and by adding and subtracting $w\bar{\mathbf{x}}\boldsymbol{\delta}$ and $w\bar{h}$, we finally have that:

$$\begin{aligned}y = \mu_0 + w\cdot(\mu_1 - \mu_0) + \mathbf{x}\boldsymbol{\delta}_0 + w\cdot[\mathbf{x}\boldsymbol{\delta}_1 - \mathbf{x}\boldsymbol{\delta}_0] + w\cdot h(t) + e_0 + w\cdot(e_1 - e_0) + \\ (w\bar{\mathbf{x}}\boldsymbol{\delta} - w\bar{\mathbf{x}}\boldsymbol{\delta}) + (w\bar{h} - w\bar{h})\end{aligned} \tag{A2.2}$$

that is:

$$y_i = \mu_0 + w_i\cdot\text{ATE} + \mathbf{x}_i\boldsymbol{\delta}_0 + w_i\cdot(\mathbf{x}_i - \bar{\mathbf{x}})\boldsymbol{\delta} + w_i\cdot(h(t_i) - \bar{h}) + \eta_i \tag{A2.3}$$

where $\eta_i = e_{0i} + w_i\cdot(e_{1i} - e_{0i})$. □

A3. Proof of proposition 3. *Ordinary Least Squares (OLS) consistency*.

Under assumption 1 and 2 (CMI), the error tem of regression (12) has zero mean conditional on (w_i, $\mathbf{x}_i$, t_i), i.e.:

$$\begin{aligned}
&\mathrm{E}\left(\eta_i \left| w_i, t_i, \mathbf{x}_i \right.\right) = \mathrm{E}\left(e_{0i} + w_i \cdot (e_{1i} - e_{0i}) \left| w_i, t_i, \mathbf{x}_i \right.\right) = \\
&\mathrm{E}\left(e_{0i} \left| w_i, t_i, \mathbf{x}_i \right.\right) + \mathrm{E}\left(w_i \cdot (e_{1i} - e_{0i}) \left| w_i, t_i, \mathbf{x}_i \right.\right) = \\
&\mathrm{E}\left(e_{0i} \left| w_i, t_i, \mathbf{x}_i \right.\right) + w_i \cdot \left(\mathrm{E}(e_{1i} - e_{0i}) \left| w_i, t_i, \mathbf{x}_i \right.\right) = \\
&\mathrm{E}(e_{0i} \left| w_i, t_i, \mathbf{x}_i \right.) + w_i \cdot \mathrm{E}(e_{1i} \left| w_i, t_i, \mathbf{x}_i \right.) - w_i \cdot \mathrm{E}(e_{0i} \left| w_i, t_i, \mathbf{x}_i \right.) = \\
&\mathrm{E}(e_{0i} \left| \mathbf{x}_i \right.) + w_i \cdot \mathrm{E}(e_{1i} \left| \mathbf{x}_i \right.) - w_i \cdot \mathrm{E}(e_{0i} \left| \mathbf{x}_i \right.) = 0
\end{aligned} \tag{A3.1}$$

The fourth equality comes from CMI; whereas, the last equality derives from the assumption of exogeneity of **x**. □

A4. Proof of proposition 4. *Analytical standard error for the dose-response function*.

Formula (18) 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. Indeed, when w=1, Eq. (17) can be written as:

$$\widehat{\mathrm{ATE}}(t_i) = \widehat{\mathrm{ATET}} + \hat{a} \cdot T_{1i} + \hat{b} \cdot T_{2i} + \hat{c} \cdot T_{3i} \tag{A4.1}$$

This implies that:

$$\begin{aligned}
&Var\left\{\widehat{\mathrm{ATE}}(t_i)\right\} = Var\left\{\hat{a} \cdot T_{1i} + \hat{b} \cdot T_{2i} + \hat{c} \cdot T_{3i}\right\} = \\
&Var\left\{\hat{a} \cdot T_{1i}\right\} + Var\left\{\hat{b} \cdot T_{2i}\right\} + Var\left\{\hat{c} \cdot T_{3i}\right\} + 2Cov\left\{\hat{a} \cdot T_{1i}; \hat{b} \cdot T_{2i}\right\} + \\
&2Cov\left\{\hat{a} \cdot T_{1i}; \hat{c} \cdot T_{3i}\right\} + 2Cov\left\{\hat{b} \cdot T_{2i}; \hat{c} \cdot T_{3i}\right\} = \\
&T_{1i}^2 \hat{\sigma}_a^2 + T_{2i}^2 \hat{\sigma}_b^2 + T_{3i}^2 \hat{\sigma}_c^2 + 2T_{1i}T_{2i}\hat{\sigma}_{a,b} + 2T_{1i}T_{3i}\hat{\sigma}_{a,c} + 2T_{2i}T_{3i}\hat{\sigma}_{b,c}
\end{aligned} \tag{A4.2}$$

That is equivalent to Eq. (18). □

Proposition 5. *Consistent estimation of system (21)*.

We show that the IV procedure proposed in the text to tackle *w* and *t* endogeneity leads to a consistent estimation of the basic parameters in equation (21.1). We show consistency for the first step (type-2 tobit), and then for the second step (2SLS) under assumptions 4, 5 and 6 and by assuming (22), provided that $e_1=e_0$.

First step consistency. Here, we estimate equations (10.2)-(10.3) jointly by a type-2 tobit model. As said in the text, a Heckman two-step procedure (Heckman, 1979) provides a consistent estimation for such a two-equation system. For the proof, we remind directly to the original article. Observe, however, that under errors' joint normality a maximum-likelihood (ML) estimation leads to more efficient estimates. In such a specific case, the form of the likelihood function is:

$$L = \prod_{i=1}^{N} \{Pr(w_i^* \le 0)\}^{1-w_i} \{f(t_i \mid w_i^* > 0) \times Pr(w_i^* > 0)\}^{w_i} \qquad \text{(A5.1)}$$

where the first term represents the contribution when $w_{1i}^* \le 0$, since $w_{1i} = 0$; the second term is the contribution in the opposite case, i.e. $w_{1i}^* > 0$. When errors are jointly normal, then $f(w^*;t')$ is bivariate normal and the conditional density in the second term is in turn univariate normal.

Second step consistency: by definition, the predicted values of w^* (i.e. $\hat{p}_{wi}$) and t_i (i.e. $\hat{t}_i$) from the previous type-2 tobit estimation are the orthogonal projections of these variables on the vector space generated by all the exogenous variables of system (21). Call this vector space as Γ. As such, $\hat{p}_{wi}$ and $\hat{t}_i$ have three properties: (i) as functions of exogenous variables, they are exogenous themselves; (ii) as orthogonal projections, they provide the best representation of w_i^* and t_i on Γ; (iii) they are correlated with w_i^* and t_i,

respectively. Given these properties, $\hat{p}_{wi}$ and $\hat{t}_i$ are good candidates to be instruments when w_i and t_i are endogenous in equation (21.1).

At this point, one would be tempted to replace w and t with $\hat{p}_{wi}$ and $\hat{t}_i$ directly into equation (21.1), and then perform an OLS regression of this equation: this would be a correct IV procedure leading to consistent estimation only if the type-2 tobit model is "correctly" specified.

When the type-2 tobit is not correctly specified, either $\hat{p}_{wi}$ or $\hat{t}_i$ are affected by a measurement error, implying that the second-step OLS of the previous IV approach would provide inconsistent estimation of parameters' in (21.1). To overcome this problem, as usual in the presence of measurement error, one could perform a two-stage least squares (2SLS) for equation (21.1) using as instruments the following exogenous variables $(\mathbf{x}_i, \hat{p}_{wi}, \hat{p}_{wi}[\mathbf{x}_i - \overline{\mathbf{x}}], \hat{p}_{wi}\hat{T}_{1i}, \hat{p}_{wi}\hat{T}_{2i}, \hat{p}_{wi}\hat{T}_{3i})$. Operationally, this implies three steps:

1. Estimate the type-2 tobit model of w and t on $\mathbf{x}$ and z, and get $\hat{p}_{wi}$ and $\hat{t}_i$, i.e. the predicted probability of w_i and the predicted dose t_i respectively.
2. Run an OLS of w and t on (1, $\mathbf{x}$, $\hat{p}_{wi}$ and $\hat{t}_i$), thus getting the new fitted values $w_{2fv,i}$ and $t_{2fv,i}$.
3. Run a second OLS of y on $\mathbf{x}_i, w_{2\,fv,i}, w_{2\,fv,i}[\mathbf{x}_i - \overline{\mathbf{x}}], w_{2\,fv,i}\hat{T}_{1\,fv,i}, w_{2\,fv,i}\hat{T}_{2\,fv,i}, w_{2\,fv,i}\hat{T}_{3\,fv,i}$.

This 2SLS approach provides consistent estimation of the basic coefficients μ_0, $\boldsymbol{\delta}_0$, ATE, $\boldsymbol{\delta}$, a, b, c also in presence of measurement error due to type-2 tobit misspecification (see Wooldridge, 2010, pp. 937-951 for a similar approach). Of course, when the specification error is not present, then the "direct" IV approach will be more

efficient and should be employed. However, the 2SLS procedure will be more robust. See Cerulli (2014) for a Monte Carlo comparison of such models. □

Table 1. Variables used in the specification of the outcome regression model.

Covariates	
Treatment	Binary treatment indicator taking value 1 for supported and 0 for non-supported firms
Treatment level or dose	Intensity of treatment varying between 0 and 100
N. of employees	Number of company employees, as proxy of company size
Debt	Stock of company total stock of debt (long, medium and short term) on total turnover
Cash-flow	Rate of profitability, as proxy of company liquidity constrain
Labour-intensity	Labor cost to turnover
Capital-intensity	Stock of firm material assets to turnover, as measure of capital deepening
Knowledge Stock	Stock of firm immaterial assets, as measure of accumulated R&D experience
Group	Binary indicator taking value 1 if the company is part of a group and 0 otherwise
Export	Binary indicator taking value 1 if the company exports and 0 otherwise
Size	Six categories for company size, using the number of employees
Sector	Four categories representing Pavitt sector taxonomy
Location	Twenty Italian regions
Outcomes	
Gross R&D expenditure	Total R&D expenditure
Net R&D expenditure	Total R&D expenditure minus the public support received by the firm
Gross R&D per capita	Total R&D expenditure on total number of employees
Net R&D per capita	Net R&D expenditure on total number of employees
Gross R&D intensity	Total R&D expenditure on company turnover
Net R&D intensity	Net R&D expenditure on company turnover

Table 2. Representativeness of the final sample employed in this application.

	Sample	*Population*
Size		
0-10	8.78	3.08
11-20	12.24	29.48
21-50	18.2	33.98
51-250	50.05	25.67
251-500	6.39	4.05
>500	4.33	3.74
Location		
North	73.13	67.83
Center	16.68	18.09
South & Ilands	10.21	14.06
Sector		
Traditional	42.47	51.23
Scale intensive	15.6	18.03
Specialized suppliers	36.19	25.93
High-tech	5.74	4.8
Number of obs.	932	14,106

Note: "Population" figures are calculated using the 9th Unicredit survey (the intermediate one), by weighting companies though stratified sample weights in order to approximate the actual structure of Italian manufacturing companies with more than 10 employees.

Table 3. Some descriptive statistics for public financed company R&D share.

Number of obs.	235
Mean	40.32
Std. Dev.	33.67
Min	1
Max	100
Median	30

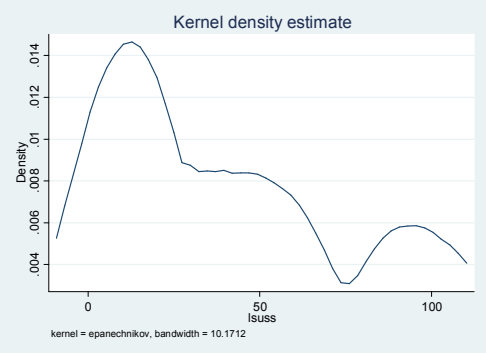


Table 4. Baseline regression for assessing the effect of public support intensity on firm R&D outcomes.

	(1) Gross R&D	(2) Net R&D	(3) Gross R&D per capita	(4) Net R&D per capita	(5) Gross R&D intensity	(6) Net R&D intensity
Treatment	308.39**	325.74**	1.66	1.22	0.01	0.01
	(142.48)	(137.25)	(1.22)	(1.19)	(0.01)	(0.01)
N. of employees	2.14***	2.11***	-0.00	-0.00	-0.00	-0.00
	(0.26)	(0.25)	(0.00)	(0.00)	(0.00)	(0.00)
Debt	-3.27*	-3.38*	0.01	0.01	0.00	0.00
	(1.95)	(1.88)	(0.02)	(0.02)	(0.00)	(0.00)
Cash-flow	-7.49**	-7.44***	0.02	0.01	0.00*	0.00*
	(2.99)	(2.88)	(0.03)	(0.02)	(0.00)	(0.00)
Labour-intensity	1.65	1.78	-0.02	-0.02	0.00	0.00
	(3.41)	(3.29)	(0.03)	(0.03)	(0.00)	(0.00)
Capital-intensity	0.96	0.95	0.01*	0.01*	-0.00	-0.00
	(0.80)	(0.77)	(0.01)	(0.01)	(0.00)	(0.00)
Knowledge Stock	4.53	4.86	0.13	0.14	0.00	0.00
	(16.85)	(16.23)	(0.14)	(0.14)	(0.00)	(0.00)
Group	159.25**	124.34*	0.86	0.66	0.01	0.00
	(69.39)	(66.84)	(0.59)	(0.58)	(0.00)	(0.00)
Export	-1.74	4.06	0.96	0.81	-0.01	-0.01
	(79.93)	(76.99)	(0.69)	(0.66)	(0.00)	(0.00)
Parameter a	-3.01	-12.27	0.01	-0.01	-0.00	-0.00
	(19.39)	(18.68)	(0.17)	(0.16)	(0.00)	(0.00)
Parameter b	0.00	0.01	-0.00	-0.00	-0.00	-0.00
	(0.48)	(0.46)	(0.00)	(0.00)	(0.00)	(0.00)
Parameter c	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	932	932	932	932	928	928
adj. R^2	0.399	0.345	-0.004	-0.011	0.041	0.034
r2	0.43	0.38	0.04	0.04	0.09	0.08
F	14.75	11.90	0.93	0.77	1.88	1.72

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Sector, size and location dummies not reported.

Table 5. Baseline regression for assessing the effect of public support intensity on firm R&D outcomes. Standardized coefficients.

	(1) Gross R&D	(2) Net R&D	(3) Gross R&D per capita	(4) Net R&D per capita	(5) Gross R&D intensity	(6) Net R&D intensity
Treatment	0.118**	0.136**	0.096	0.073	0.096	0.076
	(142.48)	(137.25)	(1.22)	(1.19)	(0.01)	(0.01)
N. of employees	0.390***	0.417***	-0.043	-0.038	-0.084	-0.082
	(0.26)	(0.25)	(0.00)	(0.00)	(0.00)	(0.00)
Debt	-0.052*	-0.059*	0.036	0.029	0.059	0.051
	(1.95)	(1.88)	(0.02)	(0.02)	(0.00)	(0.00)
Cash-flow	-0.071**	-0.076***	0.022	0.019	0.070*	0.069*
	(2.99)	(2.88)	(0.03)	(0.02)	(0.00)	(0.00)
Labour-intensity	0.015	0.018	-0.026	-0.028	0.014	0.012
	(3.41)	(3.29)	(0.03)	(0.03)	(0.00)	(0.00)
Capital-intensity	0.038	0.041	0.070*	0.076*	-0.056	-0.052
	(0.80)	(0.77)	(0.01)	(0.01)	(0.00)	(0.00)
Knowledge Stock	0.009	0.010	0.038	0.041	0.028	0.031
	(16.85)	(16.23)	(0.14)	(0.14)	(0.00)	(0.00)
Group	0.066**	0.056*	0.054	0.043	0.054	0.051
	(69.39)	(66.84)	(0.59)	(0.58)	(0.00)	(0.00)
Export	-0.001	0.002	0.051	0.045	-0.043	-0.059
	(79.93)	(76.99)	(0.69)	(0.66)	(0.00)	(0.00)
Parameter a	-0.057	-0.252	0.022	-0.034	-0.005	-0.094
	(19.39)	(18.68)	(0.17)	(0.16)	(0.00)	(0.00)
Parameter b	0.006	0.019	-0.064	-0.209	-0.156	-0.138
	(0.48)	(0.46)	(0.00)	(0.00)	(0.00)	(0.00)
Parameter c	0.012	0.090	0.056	0.142	0.187	0.133
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	932	932	932	932	928	928
adj. R^2	0.399	0.345	-0.004	-0.011	0.041	0.034
r2	0.43	0.38	0.04	0.04	0.09	0.08
F	14.75	11.90	0.93	0.77	1.88	1.72

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Sector, size and location dummies not reported.

Table 6. Description of the dataset "nlswork.dta". Data from the National Longitudinal Survey (NLS) on young women (14-26 years), based on a series of interviews carried out from 1968 to 1988.

Variable name	***Variable meaning***
idcode	NLS id
year	Interview year
birth_yr	Birth year
age	Age in current year
race	White, black, other
married	1 if married; 0 if otherwise
never_married	1 if never married; 0 otherwise
grade	Current grade completed
collgrad	1 if college graduated; 0 otherwise
south	1 if living in south, 0 otherwise
not_smsa	1 if not living in a metropolitan area (SMSA); 0 otherwise
c_city	1 if living in central city; 0 otherwise
ind_code	Type of industry
occ_code	Type of occupation
wks_ue	Weeks unemployed last year
wks_work	Weeks worked last year
msp	1 if married, with spouse present
union	1 if a union worker; 0 otherwise
wage	Hourly wage
hours	Usual number of hours worked
ttl_exp	Total work experience
tenure	Job tenure (measured in years)
wks_work	Weeks worked last year
ln_wage	ln(wage/GNP deflator)

Table 7. Heckman selection model - two-step estimates. Regression model with sample selection.

	Coefficient (standard error)
tenure3 (truncated)	
Age	3.555*
	(1.842)
age_sq	-0.036
	(0.025)
birth_yr	0.347
	(0.347)
Grade	-0.703
	(1.213)
Union	5.026***
	(1.796)
_cons	-11.42
	(25.79)
Treatment	
Age	-0.102***
	(0.032)
age_sq	0.001**
	(0.001)
birth_yr	-0.010
	(0.007)
Grade	0.070***
	(0.008)
Msp	0.051
	(0.040)
_cons	3.369***
	(0.576)
Mills	
Lambda	-103.2
	(106.9)
Rho	-1.0
Sigma	103.2
Number of observations	15344
Censored observations	557
Uncensored observations	14787
Wald chi2(5)	46.49
Prob > chi2	0.000

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

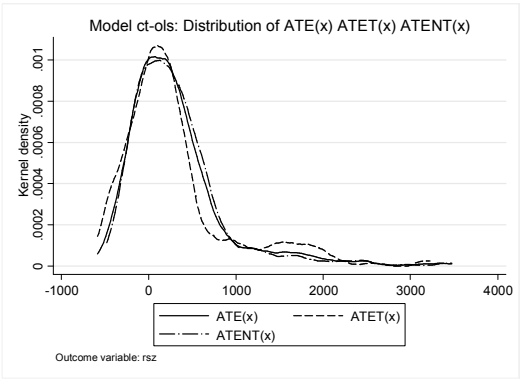
Table 8. Instrumental variables (2SLS) regression.

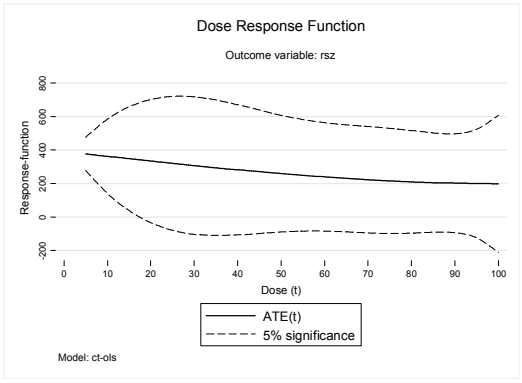
ln_wage	Coefficient (standard error)
Treatment	1.665
	(-3.191)
_ws_age	-0.667
	(0.982)
_ws_age_sq	0.013
	(0.016)
_ws_birth_yr	0.197
	(0.355)
_ws_grade	0.431*
	(0.243)
Tw_1	0.065
	(0.124)
Tw_2	0.000
	(0.002)
Tw_3	0.000
	(0.000)
Age	0.583
	(0.926)
age_sq	-0.012
	(0.015)
birth_yr	-0.190
	(0.340)
Grade	-0.356
	(0.239)
_cons	7.775
	(26.80)
Number of obs	15344
F(12, 15331)	89.53
Prob > F	0.000
R-squared	n.a.
Root MSE	0.85704
Instrumented:	
Treatment, _ws_age, _ws_age_sq, _ws_birth_yr, _ws_grade, Tw_1, Tw_2, Tw_3	
Instruments:	
age, age,_sq birth_yr, grade, probw, _ps,_age, _ps,_age_sq, _ps_birth_yr, _ps_grade, T_hatp_1, T_hatp_2, T_hatp_3	

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

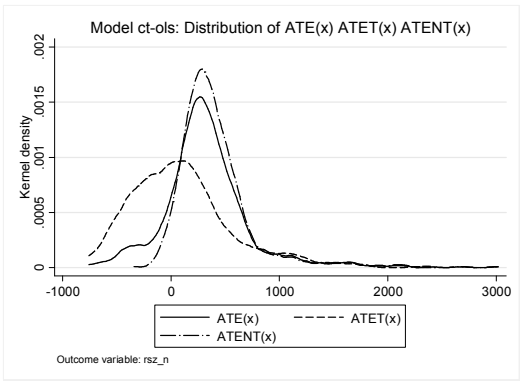
Figure 1. Distribution of ATE(**x**,t), ATET(**x**,t) and ATENT(**x**,t) and dose-response function with confidence intervals.

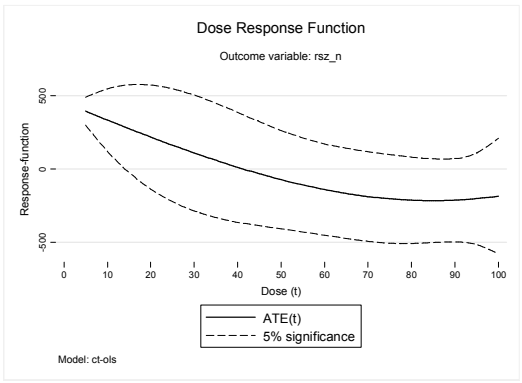
Gross R&D expenditure



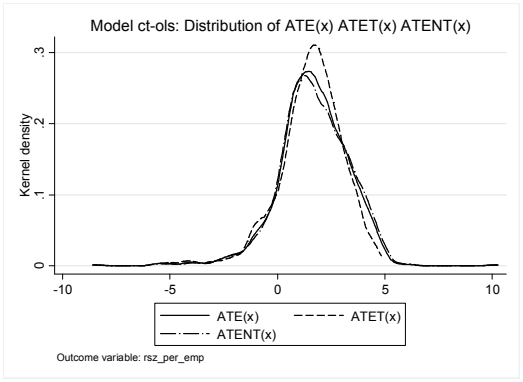


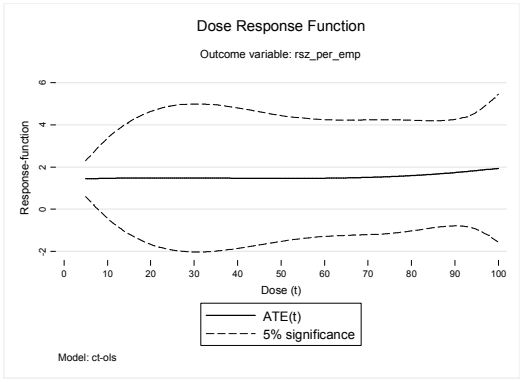
Net R&D expenditure



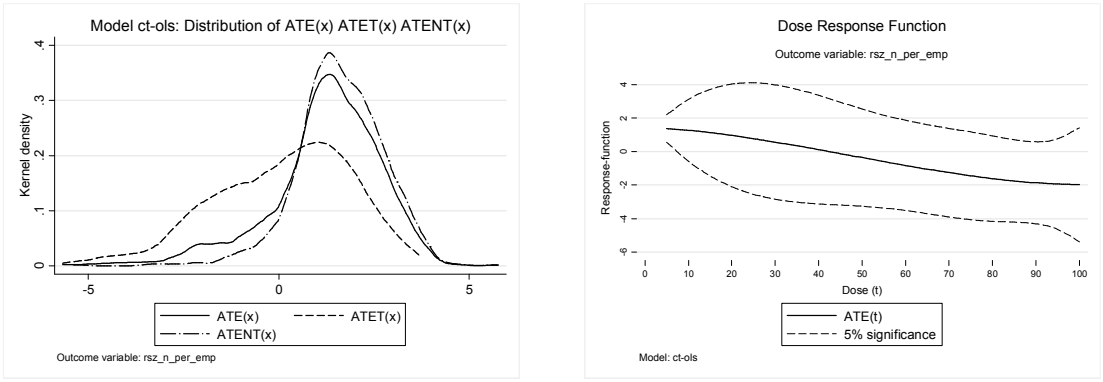


Gross R&D per capita

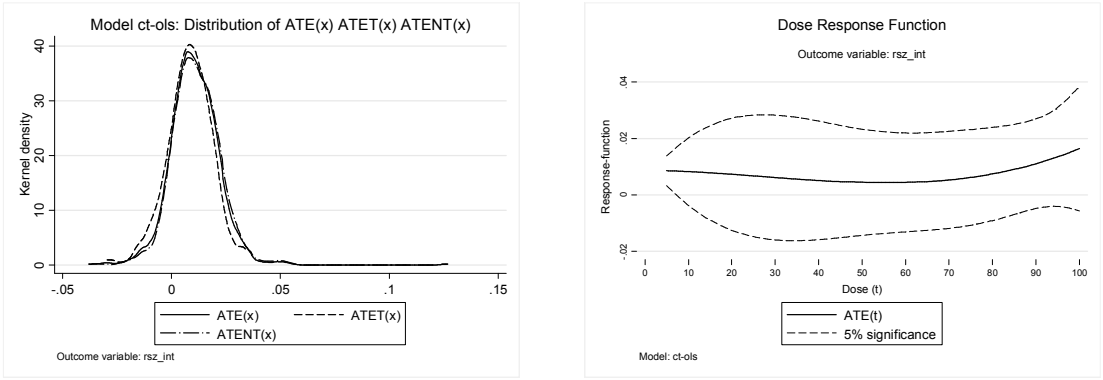




Net R&D per capita



Gross R&D intensity



Net R&D intensity

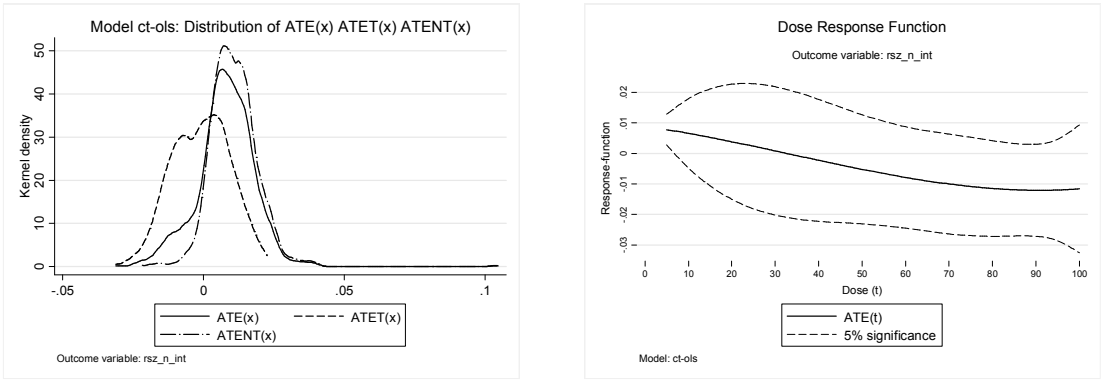
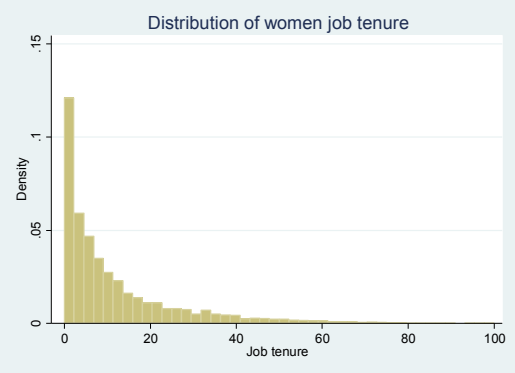
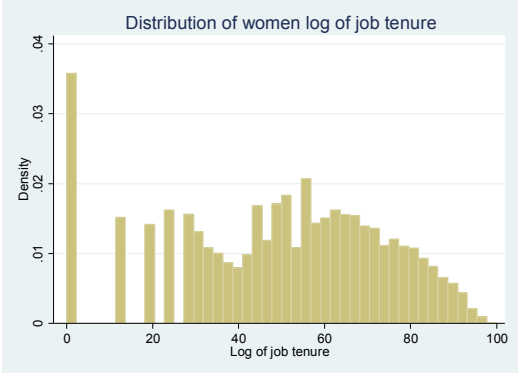


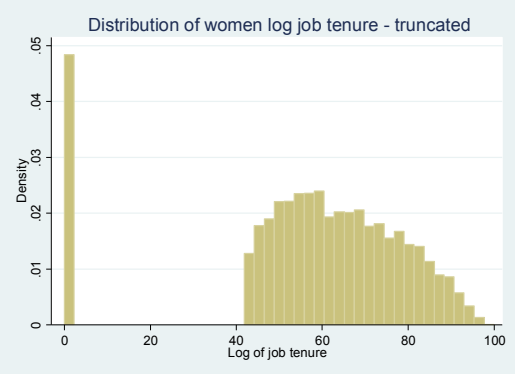
Figure 2. Distribution of "job tenure" and the "log of job tenure".


Distribution of women job tenure
Density
0
.05
.1
.15
0
20
40
60
80
100
Job tenure

(a)

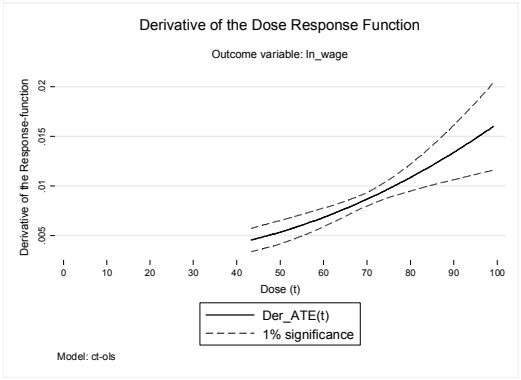

Distribution of women log of job tenure
Density
0
.01
.02
.03
.04
0
20
40
60
80
100
Log of job tenure

(b)

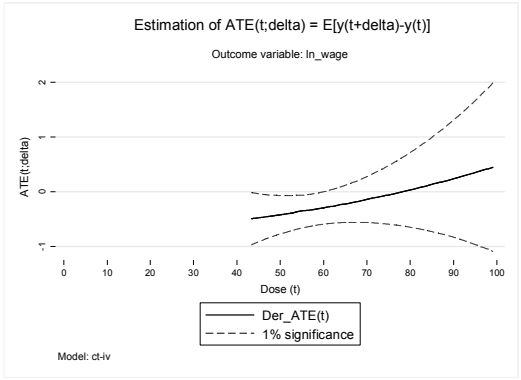

Distribution of women log job tenure - truncated
Density
0
.01
.02
.03
.04
.05
0
20
40
60
80
100
Log of job tenure

(c)

Figure 3. OLS (panel (a)) and IV (panel (b)) estimation of the dose-response function of "log of job tenure" on "log of wage" by a truncation of the "log of job tenure" carried out at dose 42.



(a)



(b)