Assessing the effect of R&D contributions: a non parametric approach based on Kernel and Spline Estimator

Michela Bia
CEPS/INSTEAD
Population & Emploi
3, Avenue de la Fonte
Esch-Sur-Alzette (L-4364),Luxembourg
E-mail: michela.bia@ceps.lu

Alessandra Mattei
University of Florence
Department of Statistics "G. Parenti"
Viale Morgagni, 59
Florence (50134), Italy
E-mail: mattei@ds.unifi.it

Introduction

Firms' innovation policies are an important tool to support local enterprises. Technological investments are considered an efficient strategy to guarantee competitiveness both at the firm-level and for the economy as a whole (Jones, 2005; Aghion and Howitt, 2005). Research & Development (R&D) investments fall in the class of interventions expected to set up technological progress and facilitate growth in the long-run (e.g., Cerulli, 2010). Many recent studies deal with public policies designed to encourage firms' investment in innovative activities, including public measures aimed at fostering innovation by strengthening and extending patent rights (Gallini, 2002) and R&D investments (e.g., Almus and Czarnitzki, 2003). In this paper we focus on evaluating the impact of public R&D financial aids provided to the Luxembourgish enterprises in 2004 and 2005, using a panel dataset obtained by matching firms from the fourth and sixth Community Innovation Survey (CIS 2004 and CIS 2006). The CIS survey is carried out by "The Central Service for Statistics and Economic Studies of Luxembourg" (STATEC) along with CEPS/INSTEAD, and collects information about product and process innovation, and organizational and marketing innovation.

Although many aspects of the effect of public R&D subsidies have been investigated, there is a lack of empirical evidence on the analysis of research incentives provided by by the Luxembourgish Ministry of Economy. In addition, to the best of our knowledge, the existing evaluation studies on the effectiveness of R&D measures in Luxembourg focus on the effect of receiving versus not receiving R&D subsidies (e.g., Nguyen, 2007; Czarnitzki and Lopes Bento, 2010). Our contribution to the existing literature is twofold. First, we advance the evidence on the evaluation of financial measures to firms in Luxembourg. Second, a distinct feature of the present paper is that it assesses the impact of the intensity of R&D subsidies, by using the amount of policy exposure as a continuous variable. As the amount of financial aid is related to the local labour market conditions and firms' performances, we expect that firms receiving different amounts of contribution will differ in their labour market outcomes. For this reason, we argue that it is important to go beyond estimation of the causal effects of public policies employing a binary discrete intervention (to be exposed or not to a policy), and instead estimate dose-response functions and marginal treatment effect functions of receiving different levels of R&D financial aid, which also allow us to account for uncovering heterogeneities in the contribution levels.

In our empirical application, the treatment variable is the intensity of the R&D incentive in Euro

(per 1000), and the outcome variable is the innovation sales (on log scale) in 2006 linked to innovation processes set up between 2004 and 2005. One difficult with this analysis is that the amount of aid is not exogenous to the firms' characteristics, implying that firms exposed to different levels of the treatment variable can systematically differ in important ways other than the observed treatment. In order to adjust for systematic differences in background characteristics occurring between firms receiving different levels of R&D financial aid, a key identifying assumption is that selection into levels of the treatment is random conditional on a set of observable pre-treatment variables (unconfoundedness; Rosenbaum and Rubin, 1983).

Under unconfoundedness (Rosenbaum and Rubin, 1983), we use Generalized Propensity Score (GPS) methods (Hirano and Imbens, 2004; Imai and Van Dyk, 2004) to estimate average treatment effects (on the treated) of different contribution levels, by employing both parametric and semi-parametric estimators of the dose-response function. Specifically, we first estimate the GPS using a flexible parametric approach based on generalized linear models; and then we estimate the dose-response function using the estimated GPS in either parametric and non-parametric regression estimators. In the full parametric approach, the conditional expectation of the outcome variable given the treatment variable and the estimated GPS is computed using a polynomial approximation (e.g., Hirano and Imbens, 2004; Mattei and Bia, 2008). As far as the semi-parametric approach is concerned, we apply both the nonparametric inverse-weighting estimator proposed by Flores et al. (2011) and new nonparametric estimators based on spline technique. These alternative approaches will be compared and contrasted one each other by simulation. We then apply them to the data from CIS 2004-2006, to estimate the effect of different levels of exposure to R&D subisidies on firms' innovation sales.

Estimation Strategy

Using the potential outcome approach to causal inference (Rubin, 1974, 1978), we estimate a continuous dose-response function that relates each value of the dose, i.e., incentive level, to the post-treatment level of firms' innovation sales. Formally, consider a set of N enterprises, and denote each of them by subscript i: $i = 1, \ldots, N$. For each enterprise i, we observe a vector of pre-treatment variables, X_i, the received incentive amount, T_i, and the value of the outcome variable associated with this treatment level, $Y_i = Y_i(T_i)$.

In order to formally describe the econometric framework we adopt, extra notation is required. Let $Y_i(t)$ denote a random variable that maps a particular potential treatment, t, $t \in \mathcal{T} \subset \mathbb{R}$, to a potential outcome. We are interested in the average dose-response function, $\mu(t) = E[Y_i(t)]$. Following Hirano and Imbens (HI) (2004), we assume that $\{Y_i(t)\}_{t\in\mathcal{T}}$, T_i, and X_i, $i = 1, \ldots, N$ are defined on a common probability space, that T_i is continuously distributed with respect to Lebesgue measure on $\mathcal{T}$, and that $Y_i = Y_i(T_i)$ is a well defined random variable. Throughout this article, we make the Stable Unit Treatment Value Assumption (SUTVA, Rubin, 1990), which implies that there is no interference between firms and that each level of the treatment define a single outcome for each firm.

Our key identifying assumption in estimating the dose-response function is that assignment to treatment is 'weakly' unconfounded given pre-treatment variables: $Y_i(t) \perp T_i|X_i$ for all $t \in \mathcal{T}$ (weak unconfoundedness, Hirano and Imbens, 2004). The GPS is defined as the conditional density of the actual treatment given the observed covariates: $r(t, x) = f_{T|X}(t|x)$. Let $R_i = r(T_i, X_i)$ denote the conditional density at the treatment actually received. The GPS is a balancing score (e.g., Rosenbaum and Rubin, 1983), that is, within strata with the same value of $r(t, x)$, the probability that $T = t$ does not depend on the value of X. In combination with the weak unconfoundedness assumption, this balancing property implies that $f_T\left(t|r(t, X_i), Y_i(t)\right) = f_T\left(t|r(t, X_i)\right)$, for every $t \in \mathcal{T}$. As a result, the GPS can be used to eliminate any bias associated with differences in the covariates. Formally, if the

assignment to the treatment is weakly unconfounded given pre-treatment variables X_i, then

$$\beta(t,r) = E\left[Y_i(t)|r(t,X_i) = r\right] = E\left[Y_i|T_i = t, R_i = r\right]$$

and the dose-response function is

$$\mu(t) = E\left[\beta(t, r(t,X_i))\right].$$

We estimate the dose-response function using a two-step procedure. In the first stage, we estimate the GPS using a parametric but flexible approach. Let $\widehat{R}_i$ denote the estimated GPS at the treatment actually received, and let $\widehat{R}_i^t = \widehat{r}(t, X_i)$ the estimated score at a specific treatment level, t. In the second stage, we estimate the dose-response function using the estimated GPS by following two steps. The first step involves estimating the conditional expectation of Y_i given T_i and the estimated GPS $\widehat{R}_i$, $E(Y_i|T_i, \widehat{R_i})$. The second step involves averaging this conditional expectation over $\widehat{R}_i^t$ to get the value of the dose-response function at t.

In this paper, we apply both parametric and non-parametric partial mean (e.g., Newey, 1994). Following HI, we implement a parametric mean approach by assuming a (flexible) parametric form for the regression function of Y_i on T_i and $\widehat{R}_i$. Specifically,

$$E(Y_i|(T_i, \widehat{R_i}) = h(T_i, \widehat{R_i}; \alpha) \qquad \text{and} \qquad \widehat{E[Y_i(t)]} = \frac{1}{N}\sum_{i=1}^{N} h(t, \widehat{R}_i^t; \widehat{\alpha}).$$

As far as the semi-parametric approach is concerned, we apply the nonparametric Inverse-Weighting (IW) estimator based on the kernel method proposed by Flores et al. (2011), and propose new nonparametric estimators based on spline technique. Following Flores et al. (2011), we implement the IW estimator by choosing a global bandwidth based on the procedure proposed by Fan and Gijbels (1996). The unknown terms appearing in the optimal global bandwidth is estimated by employing a global polynomial of order p plus 3, where p is the order of the local polynomial fitted.

The IW Kernel Estimator of the average dose-response function is given by:

$$\widehat{E[Y_i(t)]} = \frac{D_0(t)S_2(t) - D_1(t)S_1(t)}{S_0(t)S_2(t) - S_1^2(t)}$$

where $S_j(t) = \sum_{i=1}^{N} \tilde{k}_{hX}(T_i - t)(T_i - t)^j$, $D_j(t) = \sum_{i=1}^{N} \tilde{k}_{hX}(T_i - t)(T_i - t)^j Y_i$, and $\tilde{k}_{h_X}(T_i - t) = K_h(T_i - t)/\widehat{R_i^t}$.

Our spline estimator of the average dose-response function can be defined as:

$$\widehat{E[Y_i(t)]} = \frac{1}{N}\sum_{i=1}^{N} g(t, \widehat{R}_i^t),$$

where $g(t, \widehat{R}_i^t)$ is a polynomial approximation of the conditional expectation $\beta(t, \widehat{R}_i^t)$. Specifically, $g(t, \widehat{R}_i^t)$ is a piecewise function of the form:

$$g(t, \widehat{R}_i^t) = \begin{cases} g_1(t, \widehat{R}_i^t) & \text{if} \quad k_1 \leq t < k_2 \\ g_2(t, \widehat{R}_i^t) & \text{if} \quad k_2 \leq t < k_3 \\ \qquad\vdots & \\ g_{p-1}(t, \widehat{R}_i^t) & \text{if} \quad k_{p-1} \leq t < k_p \end{cases}$$

where g_j is a pre-fixed degree polynomial and $k_1 < \ldots < k_p$ are p distinct knots $k_1 < \ldots < k_p$ in the support of T, $\mathcal{T}$. The piecewise function g must interpolate all knots and be twice continuously

differentiable on the interval $[k_1, k_p]$. In this paper, we use a natural cubic spline of the treatment variable, therefore

$$g_j(t, \widehat{R}_i^t) = a_j(t-k_k)^3 + b_j(t-k_j)^2 + c_j(t-k_j) + d_j + \delta \widehat{R}_i^t \quad \text{for } j = 1, \ldots, p-1$$

with $\partial^2/\partial t\, g_1(t, \widehat{R}_i^t) = 0$ and $\partial^2/\partial t\, g_{p-1}(t, \widehat{R}_i^t) = 0$. Since the curve $g(t, \widehat{R}_i^t)$ must be continuous across its entire interval, each sub-function must join at the knots, so $g_j(k_j, \widehat{R}_i^{k_j}) = g_{j-1}(k_j, \widehat{R}_i^{k_j})$ for $j = 2, \ldots, p-1$. Also, to make the curve smooth across the interval, the derivatives must be equal at the knots; that is, $\partial/\partial t\, g_{j-1}(k_j, \widehat{R}_i^{k_j}) = \partial/\partial t\, g_j(k_j, \widehat{R}_i^{k_j})$ and $\partial^2/\partial t\, g_{j-1}(k_j, \widehat{R}_i^{k_j}) = \partial^2/\partial t\, g_j(k_j, \widehat{R}_i^{k_j})$ for $j = 2, \ldots, p-1$.

In order to address overfitting problems, we also develop an estimator based on penalized spline:

$$g(t, \widehat{R}_i^t) + \lambda \int \partial^2/\partial t\, g(t, \widehat{R}_i^t).$$

In this paper, we focus on penalized cubic spline.

Simulations

We compare the alternative estimation strategies previously described by simulation. We generate 30 samples of size $n = 50$, where outcome (innovation sales in 2006) and treatment (public R&D contributions) are randomly generated from the reference populations. Specifically, we consider two different "true" unit-level dose-response functions, which are linear and non linear in the treatment parameter, respectively:

$$Y_i(t) = 10 + 0.005 \cdot t + 0.01 \left(\frac{t^2}{1000} \right) + 0.01 \cdot r(t, X_i) + 0.001 \cdot t \cdot r(t, X_i) + e_i$$

$$Y_i(t) = 5 + \frac{1}{890} \cdot t + \exp\left(-50 \cdot \left(\frac{t}{890 - 0.5} \right)^2 \right) + r(t, X_i) + 0.0075 \cdot r(t, X_i) \cdot t + e_i$$

where e_i is an error term normally distributed: $e_i \sim \mathrm{N}\left(0, 0.7^2\right)$, and the $r(t, X_i)$ is derived assuming the normality of the treatment variable (or of its transformation) conditional on the pre-treatment covariates. In our simulation, we assume that the logarithm of the treatment (amount of the financial aid) has a normal distribution, given the covariates: $\log(T_i)|X_i \sim \mathrm{N}\left(\beta_0 + \beta_1' X_i, \sigma^2\right)$.

As we can clearly see from Figure 1, both the IW kernel estimator and the Spline estimator fit almost perfectly the non linear true dose-response. Figure 2 shows the bias, the mean square error and the coverage of nominal 95% confidence intervals, that is, the percentage of Monte-Carlo replications for which the corresponding nominal 95% confidence intervals include the true value of the dose-response function. The IW kernel, Spline and Penalized Spline estimators have coverage rates of 95% for all treatment values considered in the simulation, except for treatment levels ranging from 10,000 to 60,000 euro and greater than 300,000 euro for the Spline estimator and for treatment values ranging from 10,000 to 100,000 euro and greater than 300,000 for the IW Kernel technique. Concerning the parametric approach, the dose-response function is barely misspecified and the coverage is poor. Moreover, the Kernel and Spline estimators seem to have lower bias and lower mean square error than the parametric-based estimator. Note that all techniques work perfectly when applied to estimating the second dose-response function (linear in the treatment parameter).

Application

In our empirical study we focus on evaluating the impact of R&D financial aids (Euro per 1000) provided to Luxembourgish enterprises in 2004 and 2005. We use a panel dataset from the fourth

Figure 1: Linear and non linear dose-esponse function by HI, Kernel, Spline and Penalized Spline method (n = 50)

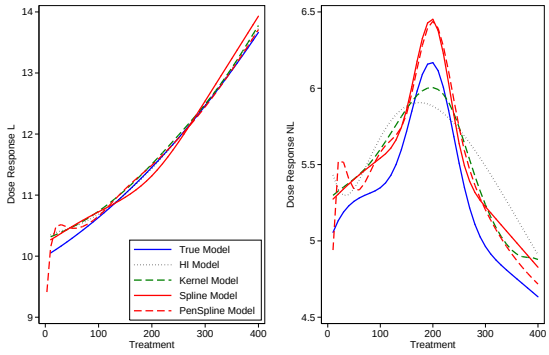
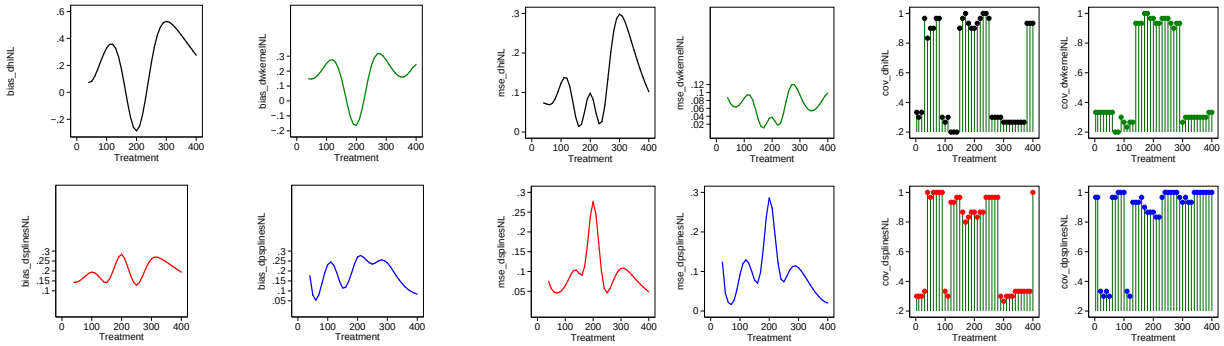


Figure 2: Bias, Mean Square Error (MSE) and Coverage Rates for the non linear dose-response function by HI, Kernel, Spline and Penalized Spline method (n = 50)



and sixth Community Innovation Survey[1]. The reference outcome is the amount of innovation sales in 2006. As many complex surveys, our study suffers from the complication that for some firms, some covariates have missing values. We deal with missing data through multiple imputation (under MAR assumption).

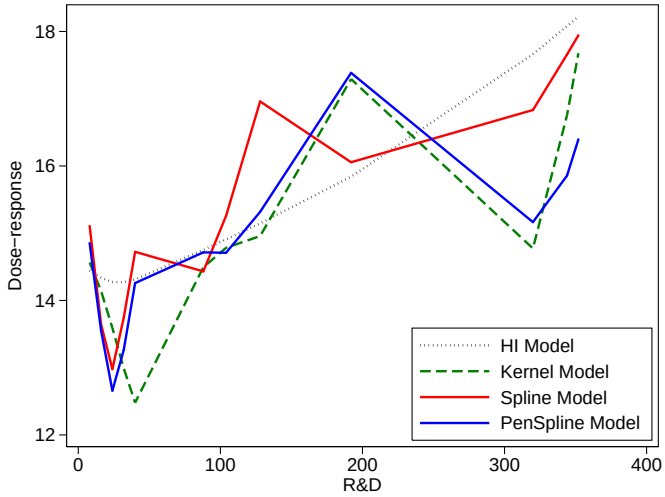
We first estimate the GPS, that is, the conditional distribution of the logarithm of the amount of R&D contribution given the covariates, and check the balancing property. Adjusting for the GPS seems to improve the balance, especially when the unadjusted differences are high (we omit these results, which are available upon request from the authors). Next, we estimate the dose-response function using the estimators previously described. The results are shown in Figure 3.

The IW Kernel and Spline estimators are very similar to each other, whereas there are important differences between these semiparametric GPS estimators and the parametric estimator. Specifically, although all estimators suggest that there exists a positive relationship between innovation sales and amount of R&D contribution, the IW Kernel and Spline estimators show a non-linear and somewhat jagged relationship, while the parametric estimator reveals a linear and smooth shape of the dose-response function.

REFERENCES (RÉFERENCES)

[1]We are particularly grateful to Vincent Dautel, who run all the programs we developed and provided, as we didn't have direct access to the data.

Figure 3: Evaluation of the amount of public R&D aids on Innovation sales (log scale)



Almus, M. and Czarnitzki, D. The effects of public R&D subsidies on firms' innovation activities: The case of eastern Germany. *Journal of Business & Economic Statistics*, **21**, 226-36, (2003).

Bia, M. and Mattei, A. A Stata package for the estimation of the Dose Response Function through adjustment for the generalized propensity score. *The Stata Journal*, **8**, 354-373, (2008).

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

Czarnitzki, D. and Lopes Bento, C. Evaluation of public R&D policies: A cross-country comparison. *CEPS/INSTEAD Working paper series* (2010).

Fan. J. and Gibels, I. *Local Polynomial Modeling and Its Applications*, Chapman and Hall: London, (1996).

Flores, C. A., Flores-Lagunes, A., Gonzalez, A. and Neumann, T.C. Estimating the effects of length of exposure to a training program: The case of Job Corps. *The Review of Economics and Statistics*, Forthcoming, (2011).

Gallini, N.T. The economics of patents: Lessons from recent U.S. patent reform. *Journal of Economic Perspectives*, **16**, 131-154, Spring.

Hirano, K. and Imbens, G. The propensity score with continuous treatment. In *Applied Bayesian Modelling and Causal Inference from Missing Data Perspectives* (eds A. Gelman and X.L. Meng), Wiley, (2004).

Jones, C.I. Growth and Ideas. In *Handbook of Economic Growth*, P. Aghion and S. Durlauf (eds.), Elsevier, 1063-1111, (2005).

Imai, K. and Van Dyk, D. A. Causal treatment with general treatment regimes: Generalizing the Propensity Score. *Journal of the American Statistical Association*, **99**(467), 854-866, (2004).

Newey, W. Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, **10**, 233-253, (1994).

Nguyen, T. U. Impact of public support on firms' innovation performance Evidence from Luxembourg's firms. *CEPS/INSTEAD Working paper series* (2007).

Rosenbaum, P. R. and Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41-55, (1983).

Rubin, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, **66**, 688-701, (1974).

Rubin, D. B. Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, **6**, 34-58, (1978).

Rubin, D. B. Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Statistical Science*, **5**, 472-480, (1990).