CEPS INSTEAD

WORKING PAPERS

A Stata package for the application of semiparametric estimators of dose-response functions

Michela BIA[1]
Carlos A. FLORES[2]
Alfonso FLORES-LAGUNES[3]
Alessandra MATTEI[4]

CEPS/INSTEAD, Luxembourg[1]
University of Miami, United-States[2]
State University of New York, United-States[3]
University of Florence, Italy[4]

Working Paper No 2013-07
April 2013

A Stata package for the application of semiparametric estimators of dose-response functions*

Michela Bia
CEPS/INSTEAD, Esch-Sur-Alzette, Luxembourg

Carlos A. Flores
Department of Economics, Coral Gables, University of Miami, Florida

Alfonso Flores-Lagunes
Department of Economics, State University of New York, Binghamton

Alessandra Mattei
Department of Statistics, Informatics, Applications "G. Parenti"
University of Florence, Italy

April, 2013

*This research is part of the "Estimation of direct and indirect causal effects using semi-parametric and non-parametric methods" project supported by an AFR grant (PDR no. 09088) from the Luxembourg 'Fonds National de la Recherche' cofunded under the Marie Curie Actions of the European Commission (FP7-COFUND). The authors would like to thank Prof. Fabrizia Mealli, Guido Imbens and Philippe Van Kerm for the insightful and useful suggestions. The Stata package is available from the authors, please contact michela.bia@ceps.lu.

Abstract

In many observational studies the treatment may not be binary or categorical, but rather continuous in nature, so focus is on estimating a continuous dose-response function. In this paper we propose a set of Stata programs to semiparametrically estimate the dose-response function of a continuous treatment, under the key assumption that adjusting for pre-treatment variables removes all biases (uncounfoundedness). We focus on kernel methods and penalized spline models, and use generalized propensity score methods under continuous treatment regimes for covariate adjustment. Several alternative parametric assumptions on the functional form of the generalized propensity score are implemented in our Stata programs, which also allow users to impose a common support condition and evaluate the balancing of the covariates using various approaches. We illustrate our routines by estimating the effect of the prize amount on subsequent labor earnings for Massachusetts lottery winners, using a data set collected by Imbens et al. (2001).

Keywords: dose-response function, generalized propensity score, kernel estimator, penalized spline estimator, weak unconfoundedness

JEL classification codes: C13 ; J31 ; J70

1 Introduction

The evaluation process in the fields of economics, sociology, law, and many other areas generally relies on the implementation of non-experimental techniques to estimate average treatment effects. Propensity score methods (Rosenbaum and Rubin, 1983) are attractive empirical tools to balance the distribution of covariates between treatment groups, and analyze the comparability of the groups in terms of observed covariates. Under the key assumption of unconfoundedness, which requires that potential outcomes are independent of the treatment conditional on the observed covariates, propensity score methods eliminate or reduce the potential bias in treatment effect estimates in observational studies. The majority of applications aim at evaluating causal effects of a binary treatment. However, in many empirical studies, treatments may take on many values, implying that participants in the study may receive different treatment levels. In such cases, the focus is on assessing the heterogeneity of treatment effects arising from variation in the amount of treatment exposure, that is, on estimating a dose-response function (DRF).

Over the last years, propensity score methods have been generalized and applied to the case of multivalued treatments (e.g., Imbens, 2000; Lechner, 2001) and, more recently, continuous treatments and arbitrary treatment regimes (e.g., Hirano and Imbens, 2004; Imai and VanDyk, 2004; Flores et al., 2012; Bia and Mattei, 2012; Kluve et al., 2012).

In this paper we build on the work by Hirano and Imbens (2004), who introduced the concept of the Generalized Propensity Score (GPS) and employed it to estimate the entire DRF of a continuous treatment. Hirano and Imbens (2004) used a *parametric* partial mean approach to estimate the DRF. Here we focus on *semiparametric* techniques. Specifically, we develop a set of Stata programs which allow users to (i) estimate the GPS under various alternative parametric assumptions; (ii) impose the common support condition as defined in Flores et al. (2012) and assess the balance of covariates after adjusting for the estimated GPS; (iii) estimate the DRF using the estimated GPS by employing either the nonparametric inverse-weighting (IW) estimator developed in Flores et al. (2012), or a new set of semiparametric estimators based on penalized spline techniques.

We illustrate these programs using a data set collected by Imbens et al. (2001). The winners of the Megabucks lottery in Massachusetts in the mid-1980's represent the reference population. We implement our programs to semiparametrically estimate the average potential post-winning labor earnings for each amount of the lottery prize (DRF). The assignment of the prize is obviously random, but unit nonresponse led to a self-selected sample where the amount of the prize received is no longer independent of background characteristics. As in the binary treatment case, the extent to which the bias generated by confounding factors is reduced heavily depends on the richness of the pre-treatment variables available in the empirical study.

The paper is organized as follows. Section 2 describes the methodological approach we refer to in the analysis. Section 3 introduces the generalized propensity score model and the semiparametric estimators of the DRF, and Section 4 shows the syntax of the `DRF.ado`. Section 6 illustrates the methods and the Stata program discussed in the paper using a survey of Massachusetts lottery winners (Imbens et al., 2001).

2 Description

We estimate a continuous DRF that relates each value of the dose (e.g., lottery prize level) to the outcome variable (e.g., post-winning labor earnings) within the potential outcome approach to causal inference (Rubin, 1974; Rubin, 1978). Formally, consider a set of N individuals, and denote each of them by subscript i: $i = 1, \ldots, N$. Under the stable unit treatment value assumption (SUTVA (Rubin, 1980; Rubin, 1990)), for each unit i there is a set of potential outcomes $\{Y_i(t)\}_{t \in \mathcal{T}}$, where $\mathcal{T}$ is a subset of the real line, $\mathcal{T} \subset \mathcal{R}$. We are interested in estimating the average DRF, $\mu(t) = E[Y_i(t)]$.

For each individual i, we observe a vector of pre-treatment covariates, X_i, the received treatment level, T_i, and the corresponding value of the outcome for this treatment level, $Y_i = Y_i(T_i)$.

The central assumption of our approach is that the assignment to treatment levels is weakly unconfounded given the set of observable variables, i.e., $Y_i(t) \perp T_i | X_i$ for all $t \in \mathcal{T}$ (Hirano and Imbens, 2004). This assumption is described as *weak unconfoundedness* because it only requires conditional independence for each potential outcome $Y_i(t)$, rather than joint independence of all potential outcomes.

Under weak unconfoundedness, we can apply the techniques based on the GPS with continuous treatments introduced by Hirano and Imbens (2004). Let $r(t, x) = f_{T|X}(t|x)$

be the conditional density of the treatment given the covariates. The GPS is defined as $R_i = r(T_i, X_i)$. 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. The weak unconfoundedness assumption, combined with the balancing score property, implies that assignment to treatment is weakly unconfounded given the GPS. Formally,

$$f_T\left(t|r(t, X_i), Y_i(t)\right) = f_T\left(t|r(t, X_i)\right).$$

for every $t \in \mathcal{T}$, Theorem 1.2.2 in Hirano and Imbens (2004). Thus, any bias associated with differences in the distribution of covariates across groups with different treatment levels can be removed using the GPS. Formally, Hirano and Imbens (2004) show that if the assignment to the treatment is weakly unconfounded given pre-treatment variables X_i, then $\mu(t) = E\left[\beta(t, r(t, X_i))\right]$, where $\beta(t, r) = E\left[Y_i(t)|r(t, X_i) = r\right] = E\left[Y_i|T_i = t, R_i = r\right]$

3 Inference

We employ two-step semiparametric estimators of the DRF. The first step involves parametrically modelling and estimating the GPS, $R(T_i, X_i)$, and assessing the common support condition and the balancing of covariates property. The second step consists of estimating the DRF, $\mu(t)$, using either the nonparametric IW kernel estimator proposed by Flores et al., (2012) or a semiparametric spline-based estimator. These two steps are implemented in the Stata routine `DRF.ado`, which is described here in detail.

3.1 Estimation of the Generalized Propensity Score

The first part of the `DRF.ado` program computes the GPS, allows users to impose an overlap condition, and tests the balancing property of the GPS.

The GPS is estimated parametrically under various alternative distributional assumptions. Specifically, we assume that

$$g(T_i|X_i) \sim \psi\left(h(\gamma, X_i), \theta\right) \tag{1}$$

where $g()$ is a link function, ψ is a probability density function (pdf), $h()$ is a flexible function of the covariates depending on an unknown parameter vector γ, and θ is a scale parameter. In the `DRF.ado` program we consider the Normal, Inverse Gaussian, and Gamma distributions; using the identity function, the logarithm, and the power function as link functions. A two-parameter Beta distribution is also implemented to address evaluation problems where the treatment variable takes on values in the interval $(0, 1)$, representing, for instance, a proportion. Maximum likelihood methods are employed to fit these models, using the official Stata command `glm` or using the user-written package `betafit` (Buis et al., 2012)[1].

Let $\widehat{R}_i$ be the estimated GPS. An important issue in GPS applications is determining the "common support" or "overlap region". The `DRF.ado` program allows users to do this using the approach proposed by Flores et al., (2012). Specifically, the sample is first divided into K groups according to the distribution of the treatment, cutting at the $100\,k/K$th, $k = 1, \ldots, K - 1$ percentile of the treatment empirical distribution. Let Q_i be the percentile unit i belongs to. For each percentile q_k, let $\hat{R}_i^k$ be the GPS evaluated at the median level of the treatment in that percentile for unit i, which is calculated for all units. The common support region with respect to percentile q_k is obtained by comparing the support of the distribution of $\hat{R}_i^k$ for those units with $Q_i = q_k$ to that of units with $Q_i \neq q_k$. Finally, the sample is restricted to units who are comparable across all the K groups simultaneously, dropping participants whose GPS is not among the common support region for all K groups. Formally:

$$CS = \bigcap_{q=1}^{K} \{i : \hat{R}_i^q \in [max\{min_{j:Q_j=q}\hat{R}_j^q, min_{j:Q_j\neq q}\hat{R}_j^q\}, min\{max_{j:Q_j=q}\hat{R}_j^q, max_{j:Q_j\neq q}\hat{R}_j^q\}]\}$$

Similarly to applications of standard propensity score methods, in GPS applications it is crucial to evaluate how well the estimated GPS works in balancing the covariates. Several methods can be applied to test the balancing properties of the GPS. In the `DRF.ado` command, two approaches are implemented: the 'blocking on the GPS' approach and an approach based on a likelihood ratio (LR) test. The 'blocking on the GPS' approach was proposed by Hirano and Imbens (2004), and it is implemented in the `DRF.ado` routine using two-sided t-test or Bayes Factor statistics (see also Bia and Mattei, 2008). The second approach was proposed by Flores et al. (2012), who suggested to compare an *unrestricted* model for T_i including all covariates plus the GPS (up to a cubic term), with a *restricted* model that sets the coefficients of all covariates to zero, using a LR test. If the GPS sufficiently balances the covariates, then they should have little explanatory power conditional on the GPS.[2]

[1] `betafit` (version 1.0.0 at time of writing) is available from the Statistical Software Components archive (or -findit `betafit`-) and must be installed separately from DRF.

[2] An alternative approach, which is not implemented in our Stata program, was proposed by Kluve et al.

3.2 Estimation of the Dose-Response Function

In the second stage of the `DRF.ado` program, we estimate the DRF by applying spline and kernel techniques. We first describe the spline estimator and then we move to the IW kernel estimator.

As in Hirano and Imbens (2004), we employ a partial mean approach (Newey, 1994) when spline techniques are considered. Specifically, we first estimate the conditional expectation of the observed outcome Y_i given the treatment actually received, T_i, and the GPS estimated in the first stage, $\widehat{R}_i$, using bivariate penalized spline smoothing based on additive spline bases, tensor products of spline bases, or radial basis functions (Ruppert et al., 2003). Mixed models provide a representation of the penalized splines that allows smoothing to be done using mixed model methodologies and software. In our routine, we use the Stata routine `xtmixed` to fit penalized spline regressions. The average dose-response function at t is estimated by averaging the estimated regression function over the estimated score function evaluated at the specific treatment level t, i.e., $\widehat{R}_i^t \equiv \widehat{r}(t, X_i)$.

The simplest bivariate penalized spline smoothing relies on additive spline bases, which can be formally defined in our setting as follows:

$$E(Y_i|T_i, \hat{R}_i) = a_0 + a_t T_i + a_r \widehat{R}_i + \sum_{k=1}^{K^t} u_k^t (T_i - k_k^t)_+ + \sum_{k=1}^{K^r} u_k^r (\widehat{R}_i - k_k^r)_+ \qquad (2)$$

where for any number z, z_+ is equal to z if z is positive and is equal to 0 otherwise, and $k_1^t < \ldots < k_{K^t}^t$ and $k_1 < \ldots < k_{K^r}^r$ are K^t and K^r distinct knots in the support of T and the support of the estimated GPS, $\widehat{R}_i$, respectively.

The additive models have many attractive features, including their simplicity. However, an additive model may not provide a satisfactory fit, so nonadditive models including interaction terms are required. To this end, we consider tensor product bases, which are obtained by forming all pairwise products of the basis functions $1, T_i, (T_i - k_1^t) \ldots, (T_i - k_{K^t}^t)$ and

(2012). It consists of regressing each covariate on the treatment variable and comparing the significance of the coefficients for specifications with and without conditioning on the GPS.

$1, \widehat{R}_i, (\widehat{R}_i - k_1^r) \ldots, (\widehat{R}_i - k_{K^r}^t)$. Formally,

$$
\begin{aligned}
E(Y_i|T_i, \hat{R}_i) &= a_0 + a_t T_i + a_r \widehat{R}_i + \lambda T_i \widehat{R}_i + \\
&\sum_{k=1}^{K^t} u_k^t (T_i - k_k^t)_+ + \sum_{k=1}^{K^r} u_k^r (\widehat{R}_i - k_k^r)_+ + \sum_{k=1}^{K^r} v_k^r \widehat{R}_i (T_i - k_k^t)_+ + \sum_{k=1}^{K^t} v_k^t T_i (\widehat{R}_i - k_k^r)_+ + \\
&\sum_{k=1}^{K^t} \sum_{k'=1}^{K^r} v_{kk'}^{tr} (T_i - k_k^t)_+ (\widehat{R}_i - k_{k'}^r)_+
\end{aligned} \tag{3}
$$

Note that estimation problems may arise when the tensor product approach is applied, especially if the sample size is relatively small. When these problems arise, the `DRF.ado` alerts users suggesting them to adopt an additive model instead.

An alternative to tensor product splines is given by the so-called *radial basis functions*, which are basis functions of the form $C(\|(t, r)' - (k, k')'\|)$ for some univariate function C. Here we consider the following function

$$
C\left(\left\| \begin{pmatrix} t \\ r \end{pmatrix} - \begin{pmatrix} k^t \\ k^r \end{pmatrix} \right\| \right) = \left\| \begin{pmatrix} t \\ r \end{pmatrix} - \begin{pmatrix} k^t \\ k^r \end{pmatrix} \right\|^2 log \left\| \begin{pmatrix} t \\ r \end{pmatrix} - \begin{pmatrix} k^t \\ k^r \end{pmatrix} \right\|
$$

where $\| \cdot \|$ is the Euclidean norm, and we assume that

$$
E(Y_i|T_i, \hat{R}_i) = a_0 + a_t T_i + a_r \widehat{R}_i + \lambda T_i \hat{R}_i + \sum_{k=1}^{K} u_k C\left(\left\| \begin{pmatrix} T_i \\ \hat{R}_i \end{pmatrix} - \begin{pmatrix} k_k^t \\ k_k^r \end{pmatrix} \right\| \right) \tag{4}
$$

where $u_1, \cdots, u_k$ are random variables with mean 0, and variance-covariance matrix $Cov(u) = \sigma_u^2 (\Omega_k^{-\frac{1}{2}})(\Omega_k^{-\frac{1}{2}})'$, with $\Omega_k = \left[C\left(\left\| \begin{pmatrix} k_k^t \\ k_k^r \end{pmatrix} - \begin{pmatrix} k_{k'}^t \\ k_{k'}^r \end{pmatrix} \right\| \right) \right]_{1 \le k, k' \le K}$.

Given the estimated parameters of the regression function (2), (3) or (4), the average potential outcome at treatment level t is estimated by averaging over $\widehat{R}_i^t$.

Flores et al. (2012) proposed to estimate the DRF using a nonparametric Inverse-Weighting (IW) estimator based on kernel methods. In this approach, the estimated scores are used to weight observations to adjust for covariate differences. Let $K(u)$ be a kernel function with the usual properties and let h be a bandwidth satisfying $h \to 0$ and $Nh \to \infty$ as $N \to \infty$. The IW approach is implemented using a local linear regression of Y on T with weighted kernel function $\tilde{K}_{h,X}(T_i - t) = K_h(T_i - t)/\hat{R}_i^t$, where $K_h(z) = h^{-1} K(z/h)$. Formally, the IW Kernel Estimator of the average DRF is defined as follows:

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

where $S_j(t) = \sum_{i=1}^{N} \tilde{K}_{hX}(T_i - t)(T_i - t)^j$ and $D_j(t) = \sum_{i=1}^{N} \tilde{K}_{hX}(T_i - t)(T_i - t)^j Y_i$, $j = 0, 1, 2$.

We implement the IW estimator using a normal kernel. By default, the global bandwidth is selected using the procedure proposed by Fan and Gijbels (1996), which is based on estimating the unknown terms appearing in the optimal global bandwidth by employing a global polynomial of order p plus 3, where p is the order of the local polynomial fitted. However, users can also arbitrary choose an alternative global bandwidth.

4 Syntax

`DRF` *varlist* [*weight*] [*if*] [*in*] , `outcome(`*varname*`)` `treatment(`*varname*`)` `gpscore(`*newvar*`)` `cutpoints(`*varname*`)` `index(`*string*`)` `nq_gps(`*#*`)` `method(`*type*`)` [`family(`*familyname*`)` `link(`*linkname*`)` `vce(`*string*`)` `common(`*#*`)` `numoverlap(`*#*`)` `test_varlist(`*varlist*`)` `test(`*type*`)` `flag(`*#*`)` `tpoints(`*vector*`)` `npoints(`*#*`)` `npercentiles(`*#*`)` `detail` `delta(`*#*`)` `bandwidth(`*#*`)` `degree1(`*#*`)` `degree2(`*#*`)` `nknots1(`*#*`)` `nknots2(`*#*`)` `knots1(`*#*`)` `knots2(`*#*`)` `additive` `nknots(`*#*`)` `knots(`*#*`)` `standardized` `estopts(`*string*`)` `det`]

Note that, in the command `DRF.ado`, the argument *varlist* represents the observed pre-treatment variables, which are used to estimate the generalized propensity score.

5 Options

5.1 Compulsory Options

I/ `outcome(`*varname*`)` specifies that *varname* is the outcome variable.

`treatment(`*varname*`)` specifies that *varname* is the treatment variable.

`gpscore(`*newvar*`)` asks users to specify the variable name for the estimated generalized propensity score.

`method(`*type*`)` specifies the *type* of approach to be used in estimating the dose-response function. The implemented approaches are bivariate penalized splines (*type* = mtspline), bivariate penalized radial splines (*type* = radialpspline) or IW kernel (*type* = iwkernel).[3]

[3]The subroutines `mtpspline` and `radialpspline` are respectively called when estimators based on penalized splines (*type* = *mtspline*) and penalized radial splines (*type* = *radialpspline*) are used.

`cutpoints(`*varname*`)` divides the range or set of the possible treatment values, Y, into intervals within which the balancing properties of the GPS are checked using a 'blocking on the GPS' approach. `varname` is a variable indicating to which interval each observation belongs to. If flag is set to 0 (see below), this option is not compulsory.

`index(`*string*`)` specifies the representative point of the treatment variable at which the GPS has to be evaluated within each treatment interval specified in `cutpoints(`*varname*`)`. The argument *string* identifies either the mean (*string* $=$ `mean`) or a percentile (*string* $=$ `p1`, ..., `p100`) of the treatment. This is used when checking the balancing properties of the GPS employing a 'blocking on the GPS' approach, where the blocking is done using the GPS evaluated at the representative point `index(`*string*`)`. If flag is set to 0 (see below), this option is not compulsory.

`nq_gps(`*#*`)` specifies that, for each treatment interval defined in `cutpoints(`*varname*`)`, the values of the GPS evaluated at the representative point `index(`*string*`)` have to be divided into # (# $\in \{1, \ldots, 100\}$) intervals, defined by the quantiles of the GPS evaluated at the representative point `index(`*string*`)`. This is used when setting the balancing properties of the GPS employing a 'blocking on the GPS' approach. If flag is set to 0 (see below), this option is not compulsory

5.2 Uncompulsory Options

I/ Global Options

a) GPS Estimation Options

`family(`*familyname*`)` specifies the distribution used to estimate the Generalized Propensity Score. The available distributional families are Gaussian (normal) (`family(gaussian)`), Inverse Gaussian (`family(igaussian)`), Gamma (`family(gamma)`), and Beta (`family(beta)`). The default is `family(gaussian)`. The Gaussian, Inverse Gaussian and Gamma distributional families are fitted using `glm`, the Beta distribution is fitted using `betafit`. The following two options are for the `glm` command, so they can be only specified when the Gaussian, Inverse Gaussian or Gamma distribution is assumed for the treatment variable.

`link(`*linkname*`)` specifies the link function for the Gaussian, Inverse Gaussian and Gamma distributional families. The available links are `link(identity)`, link(log) and link(pow), and the default is the canonical link for the family() specified (see help for glm for further details).

vce(*vcetype*) specifies the type of standard error reported for the GPS estimation, when the Gaussian, Inverse Gaussian or Gamma distribution is assumed for the treatment variable. *vcetype* may be oim, robust, cluster, clustvar, eim, opg, bootstrap, jackknife, hac, kernel, jackknife1 (see help for glm for further details).

b) Overlap Options

common(*#*) is a flag (*#* $\in \{0, 1\}$), which restricts the inference to the subsample satisfying the common support condition when it is switched on (*#* $= 1$). The default value is 1.

numoverlap(*#*) specifies that the common support condition is imposed by dividing the sample into # groups according to # quantiles of the treatment distribution. By default the sample is divided into five groups, cutting at the 20th, 40th, 60th and 80th percentiles of the distribution if common(1).

c) Balancing Property Assessment Options

test_varlist(*varlist*) specifies that the balancing property has to be assessed for each variable in *varlist*. The default *test_varlist* consists of all the variables used to estimate the generalized propensity score.

test(*type*) allows users to specify whether the balancing property is to be assessed using a 'blocking on the GPS' approach, based on either standard two-sided t-tests (test(*t_test*)) or Bayes-factors (test(*Bayes_factor*)), and/or a model comparison approach based on a likelihood ratio (LR) test (test(*L_like*)).
The 'blocking on the GPS' approach, based on standard two-sided t-tests, provides the values of the test statistics before and after adjusting for the GPS for each pre-treatment variable included in test_varlist(*varlist*) and for each pre-fixed treatment interval specified in cutpoints. Specifically, let p be the number of control variables in test_varlist(*varlist*), and let K be the number of treatment intervals specified in cutpoints(*varname*). Then, the program calculates and shows pxK values of the test statistic before and after adjusting for the GPS, where the adjustment is done by dividing (or blocking) the values of the GPS eval-

uated at the representative point `index(`*string*`)` into the number of intervals specified in `nq_gps(`*#*`)`. See Hirano and Imbens (2004) for further details.

The model comparison approach uses a likelihood ratio test to compare an unrestricted model for T_i including all the covariates plus the GPS (up to a cubic term), with a restricted model that sets the coefficients of all covariates to 0. By default, both the 'blocking on the GPS' approach and the model comparison approach are applied.

`flag(`*#*`)` allows the user to specify that `DRF.ado` estimates the GPS without performing the balancing test. The default # is 1, meaning that the balancing property must be assessed.

c) DRF Options

`tpoints(`*vector*`)` indicates that the average potential outcome function or DRF is to be estimated at each level of the treatment in *vector*. By default, the `DRF.ado` creates a vector with jth element equal to the jth observed treatment value. This option can not be used along with either the option `npoints(`*#*`)` or `npercentiles(`*#*`)` (see below).

`npoints(`*#*`)` indicates that the DRF is to be estimated at each level of the treatment belonging to a set of evenly spaced values $t_0, t_1, \ldots, t_{\#}$, that cover the range of the observed treatment. This option can not be used along with either the option `tpoints(`*#*`)` (see above) or `npercentiles(`*#*`)` (see below) .

`npercentiles(`*#*`)` indicates that the DRF is to be estimated at each level of the treatment belonging to a set of evenly spaced number of quantiles $t_{q0}, t_{q1}, \ldots, t_{q\#}$, that cover the range of the observed treatment. This option can not be used along with either the option `tpoints(`*#*`)` or `npoints(`*#*`)` (see above)

`det` displays more detailed output for the DRF estimation. When this option is not specified, the program only displays the name of the chosen technique: `method(`*radialpspline*`)`, `method(`*mtpspline*`)`, or `method(`*iwkernel*`)`).

`delta(`*#*`)` specifies that `DRF.ado` also estimates the treatment effect function $\mu(t + \#) - \mu(t)$. The default # is 0, meaning that `DRF.ado` only estimates the dose-response function, $\mu(t)$.

II/ Options for the IW Kernel estimator (`iwkernel`) bandwith(#) specifies the bandwidth to be used. By default, the global bandwidth is chosen using the automatic procedure described in Fan and Gijbels (1996). This procedure is based on estimating the unknown terms appearing in the optimal global bandwidth by employing a global polynomial of order p plus 3, where p is the order of the local polynomial fitted.

III/ Options for the radial penalized spline estimator (radialpspline)

nknots(#) specifies the number # of knots to be selected in the two-dimensional space of the treatment variable and the generalized propensity score. The default choice of # is $max(20, min(\frac{n}{4}, 150))$ where n is the number of unique T_i, R_i (Ruppert et al., 2003). When this option is specified, the subroutines radialpspline and spacefill (Bia and Van Kerm, 2013) are applied.[4] This option can not be used along with the option knots(*numlist*) (see below).

knots(*numlist*) specifies the list of knots for the treatment and the GPS variable. This option can not be used along with the option nknots(#) (see above).

standardized allows users to standardize the treatment variable and the generalized propensity score before selecting the knots. The knots are defined based on the standardized variable.

IV/ Options for the tensor-product penalized spline estimator (mtpspline)

degree1(#) specifies the power of the treatment variable included in the penalized spline model. The default is degree1(1).

degree2(#) specifies the power of the generalized propensity score included in the penalized spline model. The default is degree2(1).

nknots1(#) specifies the number (#) of knots for the treatment variable. The location of the K_kth knot is defined as $\frac{k+1}{\#+2}$th sample quantile of the unique T_i, for $k = 1, \ldots, \#$. The default choice of # is $max(5, min(\frac{n}{4}, 35))$, where n is the number of unique T_i (Ruppert et al., 2003). This option can not be used along with the option knots1(*numlist*) (see below).

nknots2(#) specifies the number (#) of knots for the generalized propensity score. The location of the K_kth knot is defined as $\frac{k+1}{\#+2}$ th sample quantile of the

[4] spacefill (version 1.0.0 at time of writing) is available from the authors (or -findit spacefill-) and must be installed separately from DRF.

unique R_i, for $k = 1, \ldots, \#$. The default choice of $\#$ is $max(5, min(\frac{n}{4}, 35))$, where n is the number of unique R_i (Ruppert et al., 2003). This option can not be used along with the option `knots2(`*numlist*`)` (see below).

`knots1(`*numlist*`)` specifies the list of knots for the treatment variable. This option can not be used along with the option `nknots1(#)` (see above).

`knots2(`*numlist*`)` specifies the list of knots for the generalized propensity score. This option can not be used along with the option `nknots2(#)` (see above).

`additive` allows user to implement penalized splines based on the additive model, without including the product terms.

V/ Mutual options for the tensor-product and radial penalized spline estimators

Mutual options for the tensor-product and radial penalized spline estimators involve either the `mtpspline` subroutine or the `radialpspline` subroutine, depending on which estimator is used.

`estopts` specifies all the possible *options* allowed when running the `xtmixed` models to fit penalized spline models (see help for `estopts` for further details).

6 Example: the Lottery Dataset

We illustrate the methods and the Stata program discussed in the previous sections by re-analyzing data coming from a survey of Massachusetts lottery winners (see Imbens et al. (2001) for details on the survey). The focus is on evaluating the effect of the prize amount on future labor earnings (from social security records). This example is also considered in Hirano and Imbens (2004).

The sample we use consists of 237 individuals who won a major prize in the lottery. The outcome of interest is earnings six years after winning the lottery ("year6"), and the treatment is the prize amount ("prize"). Pre-treatment variables are age, gender, years of high school, years of college, winning year, number of tickets bought, working status after winning, and earnings s years before winning the lottery, $s = 1, 2, \ldots, 6$. To avoid results driven by outliers, we drop observations belonging to the upper 5% of the treatment variable distribution.

We estimate the DRF and the marginal treatment effect function, which represents the marginal propensity to earn out of the yearly prize money, using the `DRF` routine. We apply both penalized spline techniques and the IW kernel estimator, although details on the output from running `DRF` are only shown for the radial penalized spline estimator (`method(`*radialpspline*`)`). First, the GPS model and summary statistics of the estimated GPS are shown, and the common support is determined. The results show that 30 observations are dropped from the analysis after imposing the common support condition. Then, the balancing property is assessed. We specify the `test(`*L_like*`)` command for the balancing test, so the model comparison approach, based on the likelihood ratio (LR) test, is reported. The LR test shows that the GPS balances the covariates, since they have little explanatory power conditional on the GPS. In particular, the restricted model for T_i that excludes the covariates cannot be rejected at the usual significance levels (p-value is 0.187), whereas the restricted model that excludes the GPS is soundly rejected (p-value is 0). Finally, the DRF and the marginal treatment effect function are estimated. Note that the option `det` is specified, so details on the estimation of the DRF are shown.

Figures 1 and 2 show the dose-response functions and the marginal treatment effect functions using three alternative estimators: IW Kernel, multivariate penalized spline, radial penalized spline. Figures 3 and 4 show these functions along with pointwise 95% confidence bands. The standard errors are computed calling the `DRF` program in the `bootstrap` command.[5] Following Hirano and Imbens (2004), we report the value of these functions at $\$10,000$ increments for all values between $\$10,000$ and $\$100,000$.

```
. use "Lottery/LotteryDataSet.dta", clear
. drop if year6==.
(35 observations deleted)
. su prize, de
              Treatment variable = Prize amount
-------------------------------------------------------------
      Percentiles      Smallest
 1%       5.3558          1.139
 5%        10.05              5
10%       11.246         5.3558       Obs                 202
25%       17.034          6.844       Sum of Wgt.         202

50%      32.1835                      Mean           57.36918
                        Largest       Std. Dev.      64.84194
75%       71.642          270.1
90%       137.27         305.09       Variance       4204.477
95%       171.73         323.32       Skewness       2.821964
99%       305.09         484.79       Kurtosis       14.18278

. drop if  prize >= r(p95)
(11 observations deleted)
```

[5]The `bootstrap` command can lead to slightly different "observed" estimates, when using the spline-based models, due to a random selection of the knots values.

```
. replace year6 = year6/1000
year6 was long now double
(92 real changes made)

. mat def tp = (10\20\30\40\50\60\70\80\90\100)

. set more off

. DRF  agew  ownhs owncoll male tixbot workthen yearm1 yearm2 yearm3 yearm4 yearm5 yearm6, ///
  outcome(year6)  treatment(prize)  gpscore(gps)  test(L_like) ///
  tpoints(tp) numoverlap(3) method(radialpspline) family(gaussian) ///
  link(log) nknots(7) det delta(1)

******************************************************
Algorithm to estimate the generalized propensity score
******************************************************

Estimation of the propensity score

Iteration 0:   log likelihood = -983.63224
Iteration 1:   log likelihood = -958.61638
Iteration 2:   log likelihood = -953.76331
Iteration 3:   log likelihood = -953.73191
Iteration 4:   log likelihood = -953.73189

Generalized linear models                          No. of obs      =       191
Optimization     : ML                              Residual df     =       178
                                                   Scale parameter =   1365.58
Deviance         =  243073.1517                    (1/df) Deviance =   1365.58
Pearson          =  243073.1517                    (1/df) Pearson  =   1365.58

Variance function: V(u) = 1                        [Gaussian]
Link function    : g(u) = ln(u)                    [Log]

                                                   AIC             =  10.12285
Log likelihood   =  -953.731889                    BIC             =  242138.2

------------------------------------------------------------------------------
             |                 OIM
       prize |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        agew |   .0158337   .0053884     2.94   0.003     .0052727    .0263947
       ownhs |   .0585064   .0742126     0.79   0.430    -.0869476    .2039604
     owncoll |  -.0108263   .0389408    -0.28   0.781    -.0871488    .0654962
        male |    .361554   .1564085     2.31   0.021      .054999    .6681091
      tixbot |  -.0174202   .0188308    -0.93   0.355     -.054328    .0194875
    workthen |    .068044   .1819285     0.37   0.708    -.2885292    .4246172
      yearm1 |  -.0033454   .0102149    -0.33   0.743    -.0233662    .0166754
      yearm2 |   .0018299   .0151926     0.12   0.904    -.0279471    .0316069
      yearm3 |  -.0190244   .0134829    -1.41   0.158    -.0454505    .0074016
      yearm4 |   .0451296   .0194034     2.33   0.020     .0070996    .0831596
      yearm5 |  -.0094795   .0147496    -0.64   0.520    -.0383882    .0194293
      yearm6 |  -.0055688   .0084792    -0.66   0.511    -.0221876    .0110501
       _cons |   2.534394    .489911     5.17   0.000     1.574186    3.494602
------------------------------------------------------------------------------
Note: The common support condition is imposed

*****************************************************************
30 observations are dropped after imposing common support
*****************************************************************
                              gps
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .0000774       .0000308
 5%     .0012373       .0000774
10%      .003318       .0003464       Obs                 161
25%     .0076905       .0004499       Sum of Wgt.         161

50%     .0092542                      Mean           .0081786
                        Largest       Std. Dev.      .0029689
75%     .0103334       .0107928
90%     .0107199        .010793       Variance       8.81e-06
95%     .0107774       .0107953       Skewness      -1.385411
99%     .0107953       .0107956       Kurtosis       3.790638

*******************************************
End of the algorithm to estimate the gpscore
*******************************************

*****************************************************************
Beginning of the assessment of the balancing property of the GPS
*****************************************************************
```

```
********************************************************
Log-Likelihood test for Unrestricted and Restricted Model
********************************************************

********************
 Unrestricted Model
********************
Iteration 0:   log likelihood = -729.44776
Iteration 1:   log likelihood = -707.12056
Iteration 2:   log likelihood = -702.86233
Iteration 3:   log likelihood = -702.85896
Iteration 4:   log likelihood = -702.85896

Generalized linear models                         No. of obs      =        161
Optimization     : ML                             Residual df     =        145
                                                  Scale parameter =    402.607
Deviance         =  58378.00995                   (1/df) Deviance =    402.607
Pearson          =  58378.00995                   (1/df) Pearson  =    402.607

Variance function: V(u) = 1                       [Gaussian]
Link function    : g(u) = ln(u)                   [Log]

                                                  AIC             =   8.929925
Log likelihood   = -702.8589642                   BIC             =   57641.21
```

prize	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
gps	-220.8653	104.513	-2.11	0.035	-425.707	-16.02369
__0000G9	-30660.92	23793.14	-1.29	0.198	-77294.61	15972.77
__0000GA	3475308	1464036	2.37	0.018	605850.8	6344765
agew	.0084917	.0037021	2.29	0.022	.0012357	.0157477
ownhs	.0003809	.0351644	0.01	0.991	-.0685401	.0693019
owncoll	.0279188	.0293008	0.95	0.341	-.0295097	.0853473
male	-.0500428	.0975023	-0.51	0.608	-.2411438	.1410582
tixbot	-.013928	.0112878	-1.23	0.217	-.0360518	.0081957
workthen	-.0463551	.1273178	-0.36	0.716	-.2958933	.2031832
yearm1	.0059355	.0082222	0.72	0.470	-.0101797	.0220506
yearm2	-.0126701	.0127231	-1.00	0.319	-.037607	.0122668
yearm3	.026964	.0152109	1.77	0.076	-.0028488	.0567768
yearm4	-.0053107	.0110529	-0.48	0.631	-.026974	.0163526
yearm5	-.0121311	.0125158	-0.97	0.332	-.0366618	.0123995
yearm6	.0013759	.0074775	0.18	0.854	-.0132798	.0160316
_cons	4.789697	.2886318	16.59	0.000	4.223989	5.355404

```
****************************************************
 Restricted Model: where the variables are excluded
****************************************************
Iteration 0:   log likelihood = -729.67192
Iteration 1:   log likelihood = -713.08728
Iteration 2:   log likelihood = -710.90557
Iteration 3:   log likelihood = -710.90514
Iteration 4:   log likelihood = -710.90514

Generalized linear models                         No. of obs      =        161
Optimization     : ML                             Residual df     =        157
                                                  Scale parameter =   410.9211
Deviance         =  64514.61459                   (1/df) Deviance =   410.9211
Pearson          =  64514.61459                   (1/df) Pearson  =   410.9211

Variance function: V(u) = 1                       [Gaussian]
Link function    : g(u) = ln(u)                   [Log]

                                                  AIC             =   8.880809
Log likelihood   = -710.9051408                   BIC             =   63716.83
```

prize	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
gps	-132.0542	84.23995	-1.57	0.117	-297.1615	33.05306
__0000G9	-44893.96	20771.74	-2.16	0.031	-85605.83	-4182.096
__0000GA	4098231	1330420	3.08	0.002	1490655	6705807
_cons	5.059741	.0710639	71.20	0.000	4.920458	5.199024

```
***************************************************
 Restricted Model: where the GPS terms are excluded
***************************************************
Iteration 0:   log likelihood = -831.02695
Iteration 1:   log likelihood = -805.81266
Iteration 2:   log likelihood = -800.34889
Iteration 3:   log likelihood = -800.33044
Iteration 4:   log likelihood = -800.33042

Generalized linear models                          No. of obs      =       161
Optimization     : ML                              Residual df     =       148
                                                   Scale parameter =  1323.861
Deviance         =  195931.4489                    (1/df) Deviance =  1323.861
Pearson          =  195931.4489                    (1/df) Pearson  =  1323.861

Variance function: V(u) = 1                        [Gaussian]
Link function    : g(u) = ln(u)                    [Log]

                                                   AIC             =  10.10348
Log likelihood   = -800.3304227                    BIC             =  195179.4

------------------------------------------------------------------------------
             |                 OIM
       prize |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        agew |   .0190611   .0079541     2.40   0.017     .0034713     .034651
       ownhs |   .0405098   .0891411     0.45   0.650    -.1342035    .2152231
     owncoll |   .0224632   .0477538     0.47   0.638    -.0711326     .116059
        male |   .3512001   .1678599     2.09   0.036     .0222008    .6801995
      tixbot |  -.0173121   .0218706    -0.79   0.429    -.0601776    .0255534
    workthen |   .1405108   .2189276     0.64   0.521    -.2885793     .569601
      yearm1 |   .0157215   .0121813     1.29   0.197    -.0081534    .0395964
      yearm2 |  -.0297081   .0267377    -1.11   0.267     -.082113    .0226969
      yearm3 |  -.0045664   .0249939    -0.18   0.855    -.0535535    .0444207
      yearm4 |   .0383519   .0274549     1.40   0.162    -.0154587    .0921626
      yearm5 |    -.01489    .019728    -0.75   0.450    -.0535561    .0237762
      yearm6 |   .0032261   .0154827     0.21   0.835    -.0271194    .0335716
       _cons |    2.33555   .6431357     3.63   0.000     1.075027    3.596073
------------------------------------------------------------------------------


Mat_LLike[11,1]
                  Lrtest
Unrestricted  -702.85896
Restricted X  -710.90514
TStatistic~X   16.092353
  p-value X    .18704265
Restrictio~X          12
Unrestricted  -702.85896
Restricted~S  -800.33042
TStatistic~S   194.94292
 p-value GPS   5.221e-42
Restrictio~S           3
           N         161

***********************************************************
 End of the assesment of the balancing property of the GPS
***********************************************************

****************
 DRF estimation
****************

Radial penalized spline estimator
Run 1  ..                                   (Cpq =         695.78)

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log restricted-likelihood = -517.16726
Iteration 1:   log restricted-likelihood = -517.11175
Iteration 2:   log restricted-likelihood = -517.11162
Iteration 3:   log restricted-likelihood = -517.11162

Computing standard errors:

Mixed-effects REML regression                   Number of obs      =       131
Group variable: _all                            Number of groups   =         1

                                                Obs per group: min =       131
                                                               avg =     131.0
                                                               max =       131
```

```
                                                 Wald chi2(2)       =      7.40
Log restricted-likelihood = -517.11162           Prob > chi2        =    0.0247

------------------------------------------------------------------------------
       year6 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       prize |  -.1037621   .0913582    -1.14   0.256    -.2828209    .0752966
         gps |    -1416.8   789.9379    -1.79   0.073     -2965.05    131.4495
       _cons |    28.2954   6.791852     4.17   0.000     14.98361    41.60718
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
_all: Identity               |
   sd(__0000I7..__0000ID)(1) |   6.37e-10   2.26e-09      5.98e-13    6.78e-07
-----------------------------+------------------------------------------------
                sd(Residual) |   13.38502    .836564      11.84184    15.12931
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =     0.00 Prob >= chibar2 = 1.0000
(1) __0000I7 __0000I8 __0000I9 __0000IA __0000IB __0000IC __0000ID

. mat li e(b)
e(b)[1,20]
            c1          c2          c3          c4          c5
y1   16.614437   13.590064   11.170396   9.5598933    8.957016
            c6          c7          c8          c9         c10
y1    9.2644672    10.19844   11.420449   12.608274    13.52113
           c11         c12         c13         c14         c15
y1  -.32016107  -.27469834  -.19155654   -.1022751 .01623526
           c16         c17         c18         c19         c20
y1  .0706448    .1149412   .12509918   .10668906   .07023563
. drop gps
. bootstrap  "DRF  agew  ownhs owncoll male tixbot workthen yearm1 yearm2 yearm3 yearm4 //
  yearm5 yearm6, outcome(year6)  treatment(prize)  gpscore(gps)  test(L_like) //
  tpoints(tp) numoverlap(3) method(radialpspline) family(gaussian) //
  link(log) nknots(7) det delta(1)"_b, reps (50)
. command: DRF agew ownhs owncoll male tixbot workthen yearm1 yearm2 yearm3 yearm4 //
  yearm5 yearm6 , outcome(year6) treatment(prize) gpscore(gps) test(L_like) //
  tpoints(tp) numoverlap(3) method(radialpspline) family(gaussian) //
  link(log) nknots(7) det delta(1)
  statistics:   b_c1       = _b[c1]
              b_c2       = _b[c2]
              b_c3       = _b[c3]
              b_c4       = _b[c4]
              b_c5       = _b[c5]
              b_c6       = _b[c6]
              b_c7       = _b[c7]
              b_c8       = _b[c8]
              b_c9       = _b[c9]
              b_c10      = _b[c10]
              b_c11      = _b[c11]
              b_c12      = _b[c12]
              b_c13      = _b[c13]
              b_c14      = _b[c14]
              b_c15      = _b[c15]
              b_c16      = _b[c16]
              b_c17      = _b[c17]
              b_c18      = _b[c18]
              b_c19      = _b[c19]
              b_c20      = _b[c20]
note: label truncated to 80 characters
Bootstrap statistics                               Number of obs    =       191
                                                   Replications     =        50

------------------------------------------------------------------------------
Variable     |  Reps  Observed      Bias  Std. Err. [95% Conf. Interval]
-------------+----------------------------------------------------------------
        b_c1 |    30  16.61444 -1.535679  4.242904    7.936724   25.29215   (N)
             |                                        1.348377   19.68809   (P)
             |                                        7.861991   19.68809  (BC)
        b_c2 |    30  13.59006 -1.374677  5.171173    3.013828    24.1663   (N)
             |                                       -9.351303   17.83814   (P)
             |                                        2.659373   17.83814  (BC)
```

```
        b_c3 |     30    11.1704 -1.816147   6.801103  -2.739421   25.08021   (N)
             |                                         -18.86036   18.12071   (P)
             |                                         -1.620142   18.12071  (BC)
        b_c4 |     30   9.559894 -2.904836   8.931471  -8.707015    27.8268   (N)
             |                                         -29.52042   19.27784   (P)
             |                                         -1.281503   19.27784  (BC)
        b_c5 |     30   8.957016 -4.620706    11.4736  -14.50914   32.42317   (N)
             |                                         -41.01822   19.24719   (P)
             |                                         -4.842792   19.24719  (BC)
        b_c6 |     30   9.264467 -6.942414   14.29988  -19.98207   38.51101   (N)
             |                                         -51.39978   19.97105   (P)
             |                                         -9.806973   19.97105  (BC)
        b_c7 |     30   10.19844 -9.677037   17.31045  -25.20541   45.60229   (N)
             |                                         -59.86525   21.59847   (P)
             |                                         -13.77467   21.59847  (BC)
        b_c8 |     30   11.42045 -12.65516   20.59241  -30.69575   53.53665   (N)
             |                                         -68.31172   24.65939   (P)
             |                                         -16.06157   24.65939  (BC)
        b_c9 |     30   12.60827 -15.76136   24.18286  -36.85123   62.06778   (N)
             |                                         -79.07858   26.85438   (P)
             |                                         -11.88224   26.85438  (BC)
       b_c10 |     30   13.52113 -18.89168   27.97965  -43.70369   70.74595   (N)
             |                                         -92.27587   28.01503   (P)
             |                                         -15.73032   28.01503  (BC)
       b_c11 |     30 -.3201611   .0299988   .2533333  -.8382858   .1979637   (N)
             |                                         -1.054751   .1862844   (P)
             |                                         -.6223834   .1862844  (BC)
       b_c12 |     30 -.2746983  -.0115712   .2310252  -.7471979   .1978012   (N)
             |                                         -1.058791   .2192596   (P)
             |                                         -.5139831   .2192596  (BC)
       b_c13 |     30 -.1915565  -.0866501   .2101515  -.6213645   .2382515   (N)
             |                                         -.8595198   .1877503   (P)
             |                                         -.3297973   .1877503  (BC)
       b_c14 |     30 -.1022751  -.1510106   .2708593  -.6562446   .4516944   (N)
             |                                         -1.161376   .1373965   (P)
             |                                         -.3703439   .1373965  (BC)
       b_c15 |     30   .0162353 -.2296974   .3323331  -.6634623   .6959328   (N)
             |                                         -1.170888   .2759393   (P)
             |                                          -.289534   .2759393  (BC)
       b_c16 |     30   .0706448 -.2615911    .354923  -.6552542   .7965438   (N)
             |                                         -1.151995   .3282185   (P)
             |                                         -.2635896   .3282185  (BC)
       b_c17 |     30   .1149412 -.2885368   .3644928  -.6305303   .8604127   (N)
             |                                         -1.185669   .3742993   (P)
             |                                         -.2724622   .3742993  (BC)
       b_c18 |     30   .1250992 -.3070457   .3806217  -.6533597   .9035581   (N)
             |                                         -1.182581   .3550077   (P)
             |                                         -.2560649   .3550077  (BC)
       b_c19 |     30   .1066891  -.313221   .3852914  -.6813202   .8946984   (N)
             |                                         -1.202775    .286127   (P)
             |                                         -.3090771    .286127  (BC)
       b_c20 |     30   .0702356 -.3111099   .4012712  -.7504562   .8909275   (N)
             |                                         -1.487813   .2179427   (P)
             |                                         -.2892012   .2179427  (BC)
------------------------------------------------------------------------------
Note:  N   = normal
       P   = percentile
       BC  = bias-corrected
end of do-file
```

Figures 1 to 4 employ the point estimates from the `DRF` routine along with the standard errors from the bootstrap. As can be seen from the figures, the two penalized spline estimators and the IW kernel estimator lead to qualitatively similar results in this application. However, it is important to note how the point estimates of the three estimators tend to differ in our application as the treatment levels increase. This is because our

Figure 1: Estimated dose-response functions

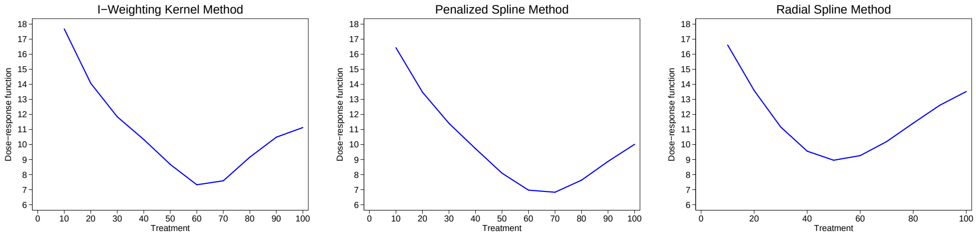
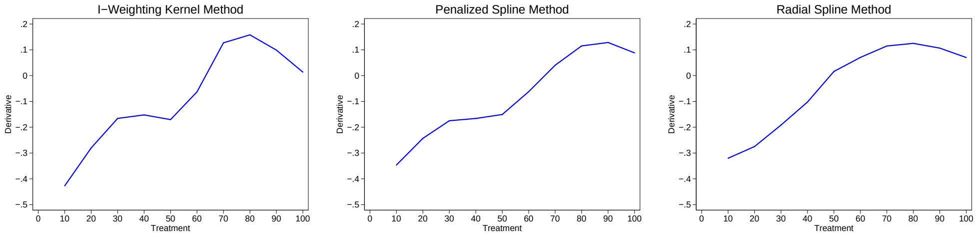


Figure 2: Estimated marginal treatment effect functions



data become scarcer as we move to higher values of the treatment.[6] Given the nonparametric nature of the methods we employ, estimation becomes noisier and the parameters are estimated less precisely in regions of the data with few observations, which is also reflected in the wider confidence intervals. In particular, note that the radial spline approach seems to be more sensitive to the size of the sample employed, as its confidence bands are wider than those of the IW and penalized splines estimators (see Figure 3), especially for values of T_i greater than 40.

[6]In particular, we do not have many observations for very low or high values (> 40) of the treatment variable.

Figure 3: 95% Confidence Bands for the dose-response functions

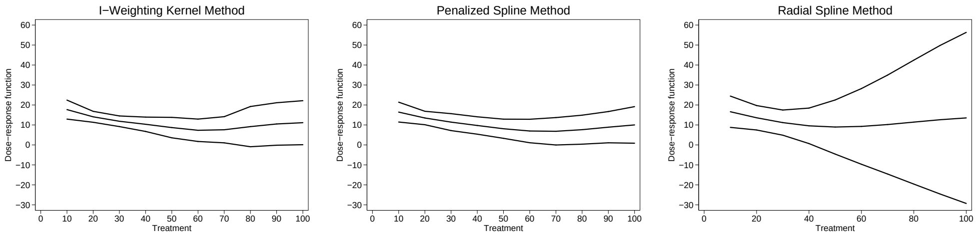
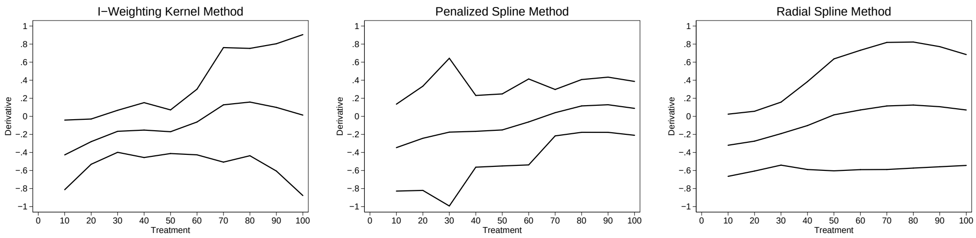


Figure 4: 95% Confidence Bands for the marginal treatment effect functions



References

[1] Bia M. and A. Mattei. Assessing the Effect of the Amount of Financial Aids to Piedmont Firms Using the Generalized Propensity Score. *Statistical Methods and Applications*, **21**, 485–516, (2012).

[2] Bia M. and P. Van Kerm. Space filling location selection with Stata. *CEPS/INSTEAD working papers series, forthcoming*, (2013).

[3] Bia M. and A. Mattei. A Stata Package for the estimation of the doseâresponse function through adjustment for the generalized propensity score. *Stata Journal*, **8**, 354–373, (2008).

[4] Buis M. L. and N. Cox and S. P. Jenkins. BETAFIT: Stata module to fit a two-parameter beta distribution. *http://EconPapers.repec.org/RePEc:boc:bocode:s435303*, (2012).

[5] Fan J. and I. Gijbels. Local Polynomial Modelling and its Applications. *Chapman and Hall, London*, (1996).

[6] Flores C. A. and A. Flores-Lagunes and A. Gonzales and T. Neuman. Estimating the Effects of Length of Exposure to Instruction in a Training Program: The Case of Job Corps. *The Review of Economics and Statistics*, **94**, 153–171, (2012).

[7] Hirano K. and G. W. Imbens. The Propensity Score with Continuous Treatments. *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives, Andrew Gelman and Xiao-Li Meng (Eds.)*, 73-84, (2004).

[8] Imai K. and D. A. Van Dyk. The Propensity Score with Continuous Treatments. *Causal treatment with general treatment regimes: Generalizing the Propensity Score*, **99**, 854–866, (2004).

[9] Imbens G. W. The role of the propensity score in estimating dose-response functions. *Biometrika*, **87**, 706–710, (2000).

[10] Imbens G. W. and D. B. Rubin and B. Sacerdote. Estimating the effect of unearned income on labor supply, earnings, savings and consumption: evidence from a survey of lottery players. *American Economic Review*, **91**, 778–794, (2001).

[11] Kluve J. and H. Schneider and A. Uhlendorff and Z. Zhao. Evaluating continuous training programs using the Generalized Propensity Score. *Journal of the Royal Statistical Society, Series A*, **175**, 587–617, (2012).

[12] Lechner M. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. *Econometric Evaluation of Labour Market Policies, Heidelberg: Physica, M. Lechner, F.Pfeiffer (Eds.)*, 43–58, (2001).

[13] Newey. W. Kernel Estimation of Partial Means and a General Variance Estimator. *Econometric Theory*, **10**, 233–253, (1994).

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

[15] Rubin D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational psychology*, **766**, 668–701, (1974).

[16] Rubin D. B. Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*, **6**, 34–58, (1978).

[17] Rubin D. B. Bias reduction using Mahalanobis's metric matching. *Biometrics*, **36**, 295–298, (1980).

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

[19] Ruppert D. and M. Wand and R. Carroll. Semiparametric Regression. *Cambridge University Press, Cambridge, UK, edition 1*, (2003).

3, avenue de la Fonte
L-4364 Esch-sur-Alzette
Tél.: +352 58.58.55-801
www.ceps.lu