Estimation of Dose-Response Functions and Optimal Doses with a Continuous Treatment

Carlos A. Flores *
Department of Economics
University of Miami

November 2007

Abstract

This paper considers the continuous-treatment case and develops nonparametric estimators for the average dose-response function, the treatment level at which this function is maximized (location of the maximum), and the maximum value achieved by this function (size of the maximum). These parameters are identified by assuming that selection into different levels of the treatment is based on observed characteristics. The proposed nonparametric estimators of the location and size of the optimal dose are shown to be jointly asymptotically normal and uncorrelated. More generally, these estimators can be used to estimate the location and size of the maximum of a partial mean (Newey, 1994). To illustrate the utility of our approach, the techniques developed in the paper are used to estimate the turning point of the environmental Kuznets curve (EKC) for NO_x, that is, the level of per capita income at which the emissions of NO_x reach their peak and start decreasing. Finally, a Monte Carlo exercise is performed partly based on the data used in the empirical application. The results show that the nonparametric estimators of the location and size of the optimal dose developed in this paper work well in practice (especially when compared to a parametric model), in some cases even for relatively small sample sizes.

Keywords and phrases: Continuous Treatment, Nonparametric Estimation, Partial Means, Location and Size of the Maximum, Environmental Kuznets Curve.
JEL classification: C13, C14, C21.

* This paper is based on the second chapter of my Ph.D. dissertation. I would like to thank my dissertation advisor, Guido Imbens, and James Powell for their advice, support and many suggestions. I would also like to thank Max Auffhammer, David Card, Kenneth Chay, Alfonso Flores-Lagunes, Keisuke Hirano, Michael Jansson, Alejandro Moreno, Thomas Rothenberg, Paul Rudd, and seminar participants at Berkeley, Western Ontario, Columbia, Missouri, UBC, Miami, UC-Davis, Illinois at Urbana-Champaign, El Colegio de Mexico, ITAM and CIDE for valuable comments and suggestions. I also thank John List and Daniel Millimet for kindly providing me with a copy of their data. Financial support from the Consejo Nacional de Ciencia y Tecnología (CONACYT)-Mexico is acknowledged. All remaining errors are my own. E-mail: caflores@miami.edu.

1 Introduction

This paper proposes a method to estimate and carry out inference for different parameters of interest when we have a continuous dose of the treatment. In particular, it focuses on estimating three objects: the entire curve of average potential outcomes or average dose-response function, the treatment level at which that curve is maximized (location of the optimal dose), and the maximum value achieved by that curve (size of the optimal dose). Under the assumption that selection into different treatment levels is based on observable characteristics, the paper estimates these objects nonparametrically based on kernel methods and establishes asymptotic normality for the estimators.

The importance of the average dose-response is obvious from a policy perspective since it gives the average outcome for all possible values of the treatment. The location and size of the optimal dose are important when a policy maker wants to apply or recommend a particular treatment dose to a population. For example, it is of interest for an agency to know the level of training that maximizes the average net benefits of a given program; or for a health provider to have an estimate of the maternal age at which health outcomes of the newborn are optimized. These two parameters can also be interpreted as the location and size of the turning point of the dose-response function, or more generally, of a given relation of interest. These objects are relevant in many areas of economics. For example, after the work by Grossman and Krueger (1991), a large number of studies have documented an inverted U-shaped relationship between some measures of pollution and per capita income. A lot of emphasis is given in this literature to estimating the turning point of this relationship.[1] Similarly, Imbs and Wacziarg (2003) estimate the turning point for the relationship between various measures of sectoral concentration and per capita income. In the area of program evaluation, Flores-Lagunes et al. (2007) and Kluve et al. (2007) estimate inverted U-shaped dose-response functions for the effect of length of exposure to a training program on earnings and the probability of employment, respectively, and the location and size of the turning point provide valuable information.[2]

The nonparametric approach presented in this paper for estimation of optimum doses or turning points has advantages over previous approaches found in the economics literature. One approach that has been previously used is to discretize the treatment, estimate average outcomes for each group, and conclude which group is best (e.g., Royer, 2003). The problem with this approach is that often discretization is arbitrary. Moreover, confidence bounds for the best group are rarely provided. Another common approach is to assume a parametric form for the relationship between the treatment and the outcome of interest and estimate the optimal

[1] Some examples are Grossman and Krueger, 1991; Selden and Song, 1994; Cropper and Griffiths, 1994; List and Gallet, 1999; Millimet et al., 2003 among others.

[2] Flores-Lagunes et al. (2007) focus on the effects of Job Corps in U.S. and Kluve et al. (2007) study training programs in Germany. Fryges (2006) and Fryges and Wagner (2007) are two more recent examples of estimated dose-response functions with an inverted-U shape in which estimation of the turning point is relevant.

treatment or turning point from it (e.g., Grossman and Krueger, 1991; Flores-Lagunes et al., 2007 and Kluve et al., 2007). Results, however, may be quite sensitive to model specifications (e.g., Millimet et al. 2003). Finally, even when some authors use nonparametric methods for estimating turning points without controlling for additional covariates (e.g., Millimet et al., 2003; Imbs and Wacziarg, 2003), they do not provide standard errors for their estimators. Contrary to the existing literature, the nonparametric estimators developed in this paper allow controlling for additional covariates nonparametrically and are shown to have an asymptotically normal distribution that can be used to undertake statistical inference. In addition, the results from a Monte Carlo exercise show that the nonparametric estimators of the location and size of the optimal dose developed in this paper work well in practice (especially as compared to a parametric model), in some cases even for relatively small sample sizes.

In order to identify the parameters of interest, we assume that selection by individuals into different treatment levels is made based on an observed set of covariates and on unobserved components not correlated with the potential outcomes. This is a straightforward extension to the continuous treatment case of the "unconfoundedness" or "selection-on-observables" assumption commonly used in the binary-treatment literature (e.g., Firpo, 2007; Imbens, 2004; Hirano et al., 2003; Heckman et al., 1999). Under this assumption we can write the average dose-response function as a partial mean, which is an average of a regression function over some of its regressors while holding others fixed. Partial means were introduced in the econometrics literature by Newey (1994). More specifically, the average dose-response function can be written as the average over the covariates of the regression function of the outcome variable on the treatment level and the covariates. Hence, the estimators presented in this paper estimate the location and size of the maximum of a partial mean. Note that in this case we do not want to maximize the regression function of the outcome variable on the treatment level and the covariates over all regressors, but only over the treatment level after we integrate over all the covariates.

The asymptotic properties of the estimators are derived using some of the general results in Newey (1994) to analyze functionals of kernel estimators. Newey (1994) derives the asymptotic distribution of a kernel-based estimator of the partial mean. The estimator of the size of the maximum of the partial mean presented in this paper behaves asymptotically as the kernel-based estimator of the partial mean evaluated at the true location of the maximum. On the other hand, the asymptotic properties of the estimator of the location of the maximum of the partial mean are closely related to those of an estimator of the first derivative of the partial mean. Hence, the scaling factor needed to obtain asymptotic normality of the location estimator is the same as the one used for a derivative estimator of the partial mean. As a result, the kernel estimators of the location and size of the maximum of the partial mean presented in this paper are based on different bandwidths. The conditions imposed on these bandwidths are stronger than those needed for asymptotic normality of the estimators of the derivative and the level of a partial mean since, in obtaining asymptotic normality of the location and size estimators, we need convergence in probability of the Jacobian resulting from a Taylor expansion around the true value of the parameters. The paper also shows that even when controlling for covariates the scaling factors needed for asymptotic normality of the location and size estimators are the same as those needed for the estimators of the location and size of the maximum of the regression function of the outcome variable on the treatment level. This comes from the fact that the rate of convergence of partial mean estimators depends on the number of regressors that are averaged out (Newey, 1994).

This paper is organized as follows. The next section presents the parameters of interest, defines the estimators, and derives their asymptotic distribution. Section 3 presents an empirical application of the methods developed in this paper. The proposed estimators are used to estimate the turning point of the "Environmental Kuznets Curve" (EKC) for NO_x, that is, the level of per capita income at which the emissions of NO_x reach their peak and start decreasing. Section 4 reports results from a Monte Carlo exercise. The simulation design is partly based on the data used in the empirical application in order to gain insight into the behavior of the estimators in situations found in empirical research. Section 5 concludes.

2 Definition and Asymptotic Distribution of the Estimators

The model is based on the potential outcome approach (Neyman, 1923; Rubin, 1974) now widely used in the program evaluation literature.[3] Assume we have a random sample of size n from a large population. We are interested in how the units in our sample respond to different doses of some treatment with the response measured by some outcome variable Y. The treatment levels, t, take on values in a set $\mathcal{T}$, where $\mathcal{T}$ is an interval. Let $Y_i(t)$ denote the potential outcome of unit i under dose t; that is, the outcome unit i would received if exposed to treatment level t. Also, let t_i be the actual treatment dose received by unit i. For each unit, out of all possible values $Y_i(t), t \in \mathcal{T}$, only $Y_i = Y_i(t_i)$ is observed, which leads to the usual missing-data problem.[4]

In this paper, we focus on estimation of three objects:

$$\mu_0(t) = E\{Y(t)\} \qquad \text{for all } t \in \mathcal{T} \tag{1}$$

$$\alpha_0 = \arg\max_{t \in \mathcal{T}} \mu_0(t) \tag{2}$$

[3] See, for instance, the surveys by Heckman, Lalonde and Smith (1999) and Imbens (2004).

[4] As noted in Hirano, Imbens and Ridder (2003), the stable-unit-treatment-value assumption (SUTVA) is implicitly assumed in this notation. SUTVA is the assumption that the potential outcome for unit i at treatment level t is not affected either by the mechanism used to assign treatment level t or by the treatment received by other units (Rubin, 1978).

and

$$\gamma_0 = \mu_0 (\alpha_0) \tag{3}$$

The first parameter is the average dose-response function. The second and third parameters are the location and size of the optimal treatment dose, respectively.

When units are randomly assigned to different levels of the treatment, estimating α_0 and γ_0 is equivalent to estimating the location and size of $E[Y|T=t]$. Kernel estimators of the location and size of the maximum of $E[Y|T=t]$ have been previously studied in the statistics literature. Müller (1985) was the first one to analyze this type of estimators in the context of the non-random regressors model and using the Gasser-Müller estimator. Müller shows that his estimators of location and size of the peak of $E[Y|T=t]$ are asymptotically jointly normal and uncorrelated. Ziegler (2000) obtains similar results when analyzing the random regressor model and using the Nadaraya-Watson estimator.[5,6]

Unfortunately, in economics we usually do not have an experiment at hand to evaluate the effects of a given treatment. A common approach in the binary-treatment literature and a natural "next step" when analyzing the effects of a given treatment is to assume that selection into treatment is based on a given set of observed covariates (e.g., Imbens, 2004). We follow an analogous approach and assume that assignment into different levels of the treatment is unconfounded given a set of covariates X with dimension equal to k, that is, we assume that selection is based on observables.[7]

Assumption 1. $\{Y(t)\}_{t\in\mathcal{T}} \perp T|X$.

Let $g_0(t,x) = E[Y|T=t, X=x]$. Assumption 1 implies that we can write the dose-response function at a given fixed value $\bar{t} \in \mathcal{T}$ as

$$\mu_0(\bar{t}) = E\left[E\left[Y(\bar{t})|X=x\right]\right] = E\left[E\left[Y(\bar{t})|T=\bar{t}, X=x\right]\right] = E\left[g_0(\bar{t},x)\right], \tag{4}$$

where the unconfoundedness assumption is used in the second equality. Randomization of the treatment levels in an experiment controls for observed and unobserved confounders by not allowing their values to differ systematically across different treatment doses. On the

[5]Another important difference between the work by Müller (1985, 1989) and Ziegler (2000) is that the former allows the order of the bandwidths for the estimators of location and size to differ and uses the same kernel for both, while Ziegler (2000) uses a bandwidth of the same order for both estimators and allows the order of the kernel to differ. Also, the conditions in Ziegler (2000) for asymptotic normality are imposed locally in a neighborhood of the location of the maximum rather than globally on a compact set.

[6]Although the literature on estimation of the maximum of a regression function is not large, the opposite is true for the related problem of estimating the mode of a density using kernel methods (e.g., Parzen, 1962; Eddy, 1982; Romano, 1988). Also, there are other approaches in the statistics literature for estimating the location of a maximum of a regression function. Some involve algorithms detecting peaks (e.g., Heckman, 1992) and the use of extreme order statistics (e.g., Chen et al., 1996). We prefer the approach based on nonparametric estimators because its extension to the case when one needs to control for additional covariates is more natural.

[7]As in Dawid (1979), we write $X \perp Y$ to denote independence of X and Y.

other hand, in the non-experimental case and under assumption 1, we need to control for systematic differences in the observed covariates across treatment doses, and $E\left[g\left(\bar{t},x\right)\right]$ does so by averaging over them. The last term in (4) is what Newey (1994) calls a partial mean, which is an average of a regression function over some conditioning variables while holding others fixed. Thus, the estimators of the location and size of the optimal treatment dose analyzed below are also useful in the more general context of estimating the location and size of the maximum of a partial mean.

The last expression in (4) suggests calculating the dose-response function following a regression approach by first computing the regression function of the observed outcome (Y) on the observed treatment (T) and covariate values (X) and then taking its expectation over the covariates.[8] As commonly done in the partial mean literature (e.g., Newey, 1994; Hausman and Newey, 1995), let $\tau(x)$ be a fixed trimming function used to bound the denominator of $g_0(t,x)$ away from zero. Using $\tau(x)$ along with assumption 1 we redefine $\mu_0\left(\bar{t}\right)$ in (1) as

$$\mu_0\left(\bar{t}\right) = E\left[\tau(x)\, g_0\left(\bar{t},x\right)\right] \tag{5}$$

Assume we observe i.i.d. data on (y_i, t_i, x_i), $i = 1, \ldots, n$. Based on (5), we define our estimators of the parameters in (1)-(3), respectively, as

$$\widehat{\mu}_\sigma\left(\bar{t}\right) = \frac{1}{n}\sum_{i=1}^{n}\tau(x_i)\ \widehat{g}_\sigma(\bar{t},x_i) \quad \text{for all} \quad \bar{t}\in\mathcal{T} \tag{6}$$

$$\widehat{\alpha} = \arg\max_{\bar{t}\in T}\widehat{\mu}_{\sigma_1}\left(\bar{t}\right) \tag{7}$$

$$\widehat{\gamma} = \widehat{\mu}_{\sigma_2}\left(\widehat{\alpha}\right) \tag{8}$$

where $\widehat{g}_\sigma(\bar{t},x)$ is the usual multivariate Nadaraya-Watson (NW) regression estimator based on bandwidth σ, which, for a kernel function $K(u)$, is given by

$$\widehat{g}_\sigma(\bar{t},x) = \frac{\sum_{j=1}^{n} y_j K\left(\frac{\bar{t}-t_j}{\sigma}, \frac{x_1-x_{1j}}{\sigma}, \cdots, \frac{x_k-x_{kj}}{\sigma}\right)}{\sum_{j=1}^{n} K\left(\frac{\bar{t}-t_j}{\sigma}, \frac{x_1-x_{1j}}{\sigma}, \cdots, \frac{x_k-x_{kj}}{\sigma}\right)} \tag{9}$$

As previously mentioned, we allow the bandwidths for $\widehat{\alpha}$ and $\widehat{\gamma}$ to differ.

Newey (1994) derives the asymptotic distribution of the partial mean estimator in (6). To state his result we introduce some notation. Let $r = (t,x)$, s be the order of the kernel used, $f_0(r)$ be the true joint density of t and x, and $\widetilde{f}_0(x)$ be the true density of x. Newey (1994)

[8] In this continuous-treatment case one could apply other methodologies to estimate the dose-response function such as weighting by the (generalized) propensity score or matching on the covariates. These methods are described in Flores (2005) for the continuous-treatment case.

shows that under some regularity conditions (see Theorem 4.1 in Newey, 1994), and assuming the bandwidth $\sigma = \sigma(n)$ satisfies $\sigma \to 0$, $n\sigma^{2k+1}/[\ln(n)]^2 \to \infty$ and $n\sigma^{2s+1} \to 0$, as $n \to \infty$, the asymptotic distribution of (6) is given by

$$\sqrt{n\sigma}(\widehat{\mu}_\sigma(\overline{t}) - \mu_0(\overline{t})) \xrightarrow{d} \mathcal{N}(0, V_0) \tag{10}$$

where $V_0 = \left[\int\{\int K(u_1, u_2)du_2\}^2 du_1\right] \times \int f_0(\overline{t}, x)^{-1}\tau^2(x)\widetilde{f}_0^2(x)Var\left[y|r = (\overline{t}, x)\right]dx$; and u in $K(u)$ is partitioned according to $r = (\overline{t}, x)$.

As discussed in Newey (1994), since $\mu(\overline{t})$ is only a function of T its nonparametric estimators will converge faster than estimators of $E[Y|T = t, X = x]$. This can be seen from the normalizing factor in (10), which is the same as the one from a nonparametric estimator of the regression function $E[Y|T = t]$. The bandwidth conditions in this result imply undersmoothing, which is reflected in the fact that the limiting distribution is centered around zero.

We now derive the joint limiting distribution of the estimators of the location and size of the optimal dose in (7) and (8). This result is derived following an approach similar to the one in Newey (1994), and by using some of his general results on functionals of kernel estimators. Newey considers two-step estimators where the first step is a vector of kernel estimators, say $\widehat{h}(r)$, and the second step is an m-estimator that depends on $\widehat{h}(r)$. To state the conditions needed for our result we introduce some additional notation. Let $q = [1\ y]'$ and $h_0(r) = E[q|r]f_0(r) = [h_{10}(r)\quad h_{20}(r)]'$. A kernel estimator of $h_0(r)$ is $\widehat{h}(r) = \frac{1}{n}\sum_{j=1}^{n} q_j K_\sigma(r - r_j) = [\widehat{h}_1(r)\quad \widehat{h}_2(r)]'$, where $K_\sigma(u) = \sigma^{-(k+1)}K(u/\sigma)$. This is the first-step kernel estimator. We impose the following conditions.

Assumption 2. Let $K(u)$ be such that $\int K(u)du = 1$; $K(u)$ is zero outside a bounded set; $K(u)$ is twice continuously differentiable with Lipschitz derivatives; and, there is a positive integer s such that for all $j < s$, $\int K(u)[\otimes_{\ell=1}^{j} u^\ell] = 0$.

Assumption 3. There is a non-negative integer $d \geq s + 1$ and an extension of $h_0(r)$ to all of $\mathbb{R}^{k+1}$ that is bounded and continuously differentiable to order d with bounded derivatives on $\mathbb{R}^{k+1}$.

Assumption 4. $E[|y|^4] < \infty$ and $E[|y|^4|r]f_0(r)$ is bounded.

Assumptions 2-4 are standard in the literature and are useful to obtain uniform convergence rates for $\widehat{h}(r)$ (e.g., Newey and McFadden, 1994). Assumption 2 requires the use of higher order kernels, which are commonly used to center the asymptotic distribution of estimators around the true parameter values.

To write the second step m-estimator we make the following assumption.

Assumption 5. $\alpha \in \mathcal{T}$, where $\mathcal{T}$ is compact, $\mu_0(\overline{\alpha}) = E[\tau(x)E[Y|r = (\overline{\alpha}, x)]]$ is uniquely maximized at α_0; and α_0 is in the interior of $\mathcal{T}$. Also, $\partial^2\mu_0(\alpha_0)/\partial\alpha^2 \neq 0$.

Let $z = (q, r)$, $\beta = [\alpha\quad \gamma]'$, and define $m_1(z, \beta, h) = \tau(x)\partial g(\alpha, x)/\partial\alpha$, where $g(\alpha, x) = h_2(\alpha, x)/h_1(\alpha, x)$. Then, under assumption 5 we have that β_0 solves $E[m_1(z, \beta_0, h_0)] =$

$\partial\mu_0(\alpha_0)/\partial\alpha = 0$. Also, let $m_2(z,\beta,h) = \tau(x)g(\alpha,x)-\gamma$, so that $E[m_2(z,\beta_0,h_0)] = 0$. Finally, let our moment vector be given by $m(z,\beta,h) = [m_1(z,\beta,h) \quad m_2(z,\beta,h)]'$. Then, the estimators in (7) and (8) are given by the vector $\widehat{\beta}$ that solves the corresponding sample equation given by

$$\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} m_1(z,\beta,\widehat{h}_{\sigma_1}) \\ m_2(z,\beta,\widehat{h}_{\sigma_2}) \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n} m(z_i,\beta,\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2}) = 0 \tag{11}$$

where note that $\widehat{h}_{\sigma_1}$ is used in the first moment and $\widehat{h}_{\sigma_2}$ in the second one.

Our goal is to derive the asymptotic distribution of $\widehat{\beta}-\beta_0$. As usual for m-estimators, in order to derive the limiting distribution of $\widehat{\beta}-\beta_0$ we expand (11) around β_0 to obtain

$$\sqrt{n}D_n(\widehat{\beta}-\beta_0) = -\left[J_n(z_i,\beta^*,\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2},\sigma_1,\sigma_2)\right]^{-1}\sqrt{n}D_n\left[\widehat{m}_n(z_i,\beta_0,\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2})\right] \tag{12}$$

where β^* is a mean value; and we let $D_n = \begin{bmatrix} \sigma_1^{3/2} & 0 \\ 0 & \sigma_2^{1/2} \end{bmatrix}$, $J_n(z_i,\beta^*,\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2},\sigma_1,\sigma_2) = D_n(\frac{1}{n}\sum_{i=1}^{n}\partial m(z_i,\beta^*,\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2})/\partial\beta)D_n^{-1}$ and $\widehat{m}_n(z_i,\beta_0,\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2}) = \frac{1}{n}\sum_{i=1}^{n} m(z_i,\beta_0,\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2})$. Hence, the normalizing factors for $\widehat{\alpha}$ and $\widehat{\gamma}$ are given by $\sqrt{n}\sigma_1^{3/2}$ and $\sqrt{n}\sigma_2^{1/2}$, respectively.

Assumption 5 is also useful to show consistency of $\widehat{\alpha}$ which, along with uniform convergence in probability of the averages appearing in J_n, is used to show convergence in probability of J_n to some matrix J. Assumption 5 implies J is invertible. Then, asymptotic normality of $\sqrt{n}D_n(\widehat{\beta}-\beta_0)$ follows from asymptotic normality of $\sqrt{n}D_n[\widehat{m}_n(z_i,\beta_0,\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2})]$. Since in this last term the moment functions depend on the kernel estimators $\widehat{h}_{\sigma_1}$ and $\widehat{h}_{\sigma_2}$, its asymptotic distribution is derived in two steps. The first one involves a linearization around h_0, and the second entails asymptotic normality of such linearization.

To derive our main result we also make the following assumption.

Assumption 6. (i) $\tau(x)$ is bounded, continuous almost everywhere and zero except on a compact set where $f_0(t,x)$ is bounded away from zero; (ii) $\widetilde{f}_0(x)$ is bounded and continuously differentiable; (iii) $E[y|r]$ and $E[y^2|r]$ are continuously differentiable; and (iv) for some $\varepsilon > 0$, $\int \sup_{||\eta||<\varepsilon}\left[\left\{1+E\left[y^4|r=(\alpha_0+\eta,x)\right]\right\}f_0(\alpha_0+\eta,x)\right]dx < \infty$.

The dominance condition in (iv) integrates over the covariates and is used when showing asymptotic normality of the linearization of $\widehat{m}_n(z_i,\beta_0,\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2})$ in (12) around h_0. A similar assumption appears also in Newey (1994) and Hausman and Newey (1995). We now present the main result in the paper.

THEOREM 1: Suppose that (i) Assumptions 1-6 are satisfied; (ii) for a scalar u_1, $\widetilde{K}(u_1)$ is symmetric, where we partition u according to $r = [t \quad x]$ and let $\widetilde{K}(u_1) = \int K(u_1,u_2)\,du_2$; (iii) for $\ell = 1,2$ and $\sigma_\ell = \sigma_\ell(n)$, $\sigma_\ell \to 0$; $n\sigma_\ell^{k+5}/\ln(n) \to \infty$; $n\sigma_\ell^{2k+1}/[\ln(n)]^2 \to \infty$; $n\sigma_1^{2s+3} \to 0$; $n\sigma_2^{2s+1} \to 0$. If $k = 0$, also assume $n\sigma_1^6 \to \infty$. Then,

$$\sqrt{n}D_n(\widehat{\beta}-\beta_0) = \begin{pmatrix} \sqrt{n\sigma_1^3}(\widehat{\alpha}-\alpha_0) \\ \sqrt{n\sigma_2}\,(\widehat{\gamma}-\gamma_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix}\right) \tag{13}$$

with $V_1 = \left(\frac{\partial^2 \mu_0(\alpha_0)}{\partial \alpha^2}\right)^{-2} \left[\int \{\widetilde{K}_{u_1}(u_1)\}^2 du_1\right] \int \psi(\alpha_0, x)\, dx$; $V_2 = \left[\int \{\widetilde{K}(u_1)\}^2 du_1\right] \int \psi(\alpha_0, x)\, dx$; $\widetilde{K}_{u_1}(u_1) = \partial \widetilde{K}(u_1)/\partial u_1$ *and* $\psi(\alpha_0, x) = f_0(\alpha_0, x)^{-1} \tau^2(x) \widetilde{f}_0^2(x) Var[y|r = (\alpha_0, x)]$. *If* $k = 0$ *then* $V_1 = \left(\partial^2 g_0(\alpha_0)/\partial \alpha^2\right)^{-2} \varphi(\alpha_0) \int \{K_u(u)\}^2 du$ *and* $V_2 = \varphi(\alpha_0) \int \{K(u)\}^2 du$, *with* $\varphi(\alpha_0) = f_0(\alpha_0)^{-1} Var[y|T = \alpha_0]$.

PROOF. See appendix.

Assumption (ii) is used in showing that the elements of the vector $\sqrt{n} D_n[\widehat{m}_n(z_i, \beta_0, \widehat{h}_{\sigma_1}, \widehat{h}_{\sigma_2})]$ in (12) are asymptotically uncorrelated. Although this condition is sufficient but not necessary for this result, it is kept in Theorem 1 since it is satisfied by most of the kernels typically used in practice. The conditions requiring $n\sigma_\ell^{k+5}/\ln(n) \to \infty$ for $\ell = 1, 2$ are useful in showing uniform convergence in probability of $\partial^2 \widehat{\mu}_{\sigma_\ell}(\alpha)/\partial \alpha^2$ to $\partial^2 \mu_0(\alpha)/\partial \alpha^2$, which is used in showing convergence in probability of $J_n(\beta^*, \cdot)$ in (12) to J. The conditions $n\sigma_\ell^{2k+1}/[\ln(n)]^2 \to \infty$ for $\ell = 1, 2$ are used for linearization of $\sqrt{n} D_n[\widehat{m}_n(z_i, \beta_0, \widehat{h}_{\sigma_1}, \widehat{h}_{\sigma_2})]$ around h_0. The conditions involving the order of the kernel imply undersmoothing and are made to center the asymptotic distribution around the true value of β_0. Finally, the requirement that $n\sigma_1^6 \to \infty$ when $k = 0$ is made to guarantee that our result is true also in the absence of covariates (i.e., in the case when doses are randomly assigned).[9] In fact, when $k = 0$ theorem 1 is comparable to previous results by Müller (1985) and Ziegler (2000).

Note that the bandwidth used for estimation of α_0 converges to zero slower than the one for estimation of γ_0. This comes from the fact that the asymptotic distribution of $\widehat{\alpha}$ is determined by the asymptotic distribution of $\partial \widehat{\mu}_{\sigma_1}(\alpha_0)/\partial \alpha$ (see 12). In general, the conditions on the bandwidth imply that if σ_ℓ is proportional to n^{δ_ℓ} and $k > 0$, then $\delta_1 \in (\max[-1/(2k+1), -1/(k+5)], -1/(2s+3))$ and $\delta_2 \in (\max[-1/(2k+1), -1/(k+5)], -1/(2s+1))$.[10] The conditions on σ_1 and σ_2 also imply that the order of the kernel used must be $s > \max\{2 + k/2, k\}$, so higher order kernels are required if $k > 0$.[11] It is also important to note that the restrictions imposed on σ_2 and on the order of the kernel for asymptotic normality of the size estimator are stronger than those needed for asymptotic normality of the partial mean estimator in Newey (1994). This comes from the fact that in our case we need convergence in probability of $J_n(z_i, \beta^*, .)$ in (12).

The asymptotic variances in (13) are very intuitive. The asymptotic variance of the size estimator is the same as that of the partial mean estimator in (10) evaluated at α_0. As for the location estimator, its asymptotic variance equals the asymptotic variance of the kernel estimator of the first derivative of the partial mean $E\left[\tau(x)\, g\left(\bar{t}, x\right)\right]$ evaluated at α_0, divided

[9]Note that $n\sigma_1^{k+5}/\ln(n) \to \infty$ implies $n\sigma_1^6 \to \infty$ for $k \geq 1$. We use $n\sigma_1^6 \to \infty$ when showing convergence in probability of the cross-term in J_n (see equation (A.6) in the appendix).

[10]For $k = 0$, theorem 1 requires $\delta_1 \in (-1/6, -1/(2s+3))$ and $\delta_2 \in (-1/5, -1/(2s+1))$.

[11]If we were only interested on estimation of the location of the maximum, the conditions on σ_1 imply $s > k$. Hence, in this case the use of a second order kernel (which is commonly used in practice) in the presence of a single covariate is allowed. Similarly, for the case without covariates (i.e., $k = 0$) Flores (2005) shows that the conditions on σ_1 and σ_2 in the joint normality result can be weakened to allow for second order kernels.

by $\left(\partial^2\mu_0(\alpha_0)/\partial\alpha^2\right)^2$. This last term is a measure of the curvature of the partial mean at α_0. Hence, as one would expect, the greater the curvature of the partial mean at the maximum the smaller the asymptotic variance of $\widehat{\alpha}$.

We now briefly discuss the result that $\widehat{\alpha}$ and $\widehat{\gamma}$ are asymptotically uncorrelated. The cross terms of $J_n(\cdot)$ in (12) equal 0 (since $\partial m_1(z,\beta,\widehat{h}_{\sigma_1})/\partial\gamma = 0$) and $(\sigma_2/\sigma_1^3)^{1/2}\partial\widehat{\mu}_{\sigma_2}(\alpha^*)/\partial\alpha$ for some mean value α^* between $\widehat{\alpha}$ and α_0. In the appendix is shown that under our assumptions $(\sigma_2/\sigma_1^3)^{1/2}\partial\widehat{\mu}_{\sigma_2}(\alpha^*)/\partial\alpha \xrightarrow{p} 0$. Note that even if we choose σ_1 and σ_2 such that $\sigma_2/\sigma_1^3 \to C < \infty$, we still get the same result since $\partial\widehat{\mu}_{\sigma_2}(\alpha^*)/\partial\alpha \xrightarrow{p} \partial\mu_0(\alpha_0)/\partial\alpha = 0$. Now consider the asymptotic covariance of the elements in the vector $\sqrt{n}D_n[\widehat{m}_n(z_i,\beta_0,\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2})]$ in (12). This term is asymptotically equivalent to $\{\int\psi(\alpha_0,x)\,dx\}[\int\widetilde{K}_{u_1}(a_n u_1)\widetilde{K}(b_n u_1)\,du_1]$, where $\widetilde{K}(u_1)$, $\widetilde{K}_{u_1}(u_1)$ and $\psi(\alpha_0,x)$ are defined in theorem 1, and $a_n = (\sigma_2/\sigma_1)^{1/2}$, $b_n = (\sigma_1/\sigma_2)^{1/2}$. Our assumption that $\widetilde{K}(u_1)$ is symmetric implies that $\int\widetilde{K}_{u_1}(a_n u_1)\widetilde{K}(b_n u_1)\,du_1 = 0$ regardless of the limits of a_n and b_n. Moreover, note that given assumption 2, even if $\widetilde{K}(u_1)$ is not symmetric $\int\widetilde{K}_{u_1}(a_n u_1)\widetilde{K}(b_n u_1)\,du_1 \to 0$ as long as σ_1 and σ_2 are of different order.

The conclusion in theorem 1 implies that the normalizing factors used to obtain asymptotic normality of the estimators of the location and size of the optimal dose when we control for covariates using a partial mean is the same as the ones used for asymptotic normality of the location and size of the maximum of $E[Y|T=t]$. As discussed in Newey (1994), this happens because of the averaging over the covariates of the non-parametric regression of Y on T and X. On the other hand, note that for calculation of $\widehat{\alpha}$ and $\widehat{\gamma}$ we first need to estimate $g_0(t,x)$ with some precision. This may be a problem if the dimension of X is large, as is usually the case for assumption 1 to be more plausible. In this case one may need to impose some restrictions such as additivity, or make part of the model parametric. For example, one approach is the use of the generalized propensity score (GPS) introduced by Imbens (2000) and extended to the continuous treatment case in Hirano and Imbens (2004). The GPS is the conditional density of the treatment given the covariates. Analogous to the binary-treatment case, in our setting one could estimate the GPS parametrically and reduce the problem of estimating nonparametrically a regression function with $k+1$ regressors to the problem of estimating one with only two regressors: the treatment level and the GPS. Some recent applications using this approach but restricting the regression function to be a parametric function include Flores-Lagunes et al. (2007), Kluve et al. (2007) and Mitnik (2007).

Finally, note that one can use theorem 1 to obtain the asymptotic distribution of the estimators of location and size of the optimal dose even if one uses the same bandwidth σ_2 for both estimators, provided that σ_2 satisfies its corresponding assumptions in theorem 1 and the estimators are appropriately normalized ($\sqrt{n}\sigma_2^{3/2}$ for location; $\sqrt{n}\sigma_2^{1/2}$ for size). However, as our results suggest, we would prefer σ_1 to go to zero slower than σ_2 since the asymptotic behavior of $\widehat{\alpha}$ is similar to that of a kernel estimator for the first derivative of a partial mean.

3 Empirical Application: the Environmental Kuznets Curve

This section illustrates how the proposed estimators can be used in practice by analyzing the relation between emissions of nitrogen oxide (NO_x) and per capita income. Since the paper by Grossman and Krueger (1991) a large number of studies have documented an inverted U-type relationship between diverse environmental indicators and income per capita, known in this literature as the environmental Kuznets curve (EKC). A lot of emphasis is given to estimating the turning point of this curve. Estimation of EKCs and their turning points for different pollutants have been at the center of discussions on worldwide organizations such as the World Bank, World Trade Organization, and environmental organizations in general, since they raise doubt on the argument that progress invariably means more pollution.[12] In addition, turning points are also commonly used in this literature to summarize results from different studies (e.g., Stern, 1998).[13]

The typical paper in this literature uses panel data with measures of some pollutants in various locations (usually countries or cities) over time. The relation is usually specified using a location and time fixed effects model, which can be written as

$$y_{it} = \xi_i + \lambda_t + g(x_{it}) + \varepsilon_{it} \tag{14}$$

where i stands for a given location and t for time, y is an indicator of environmental degradation, x is per capita income, ξ_i and λ_t are the corresponding location and time fixed effects, and ε_{it} is a random error term. The function $g(\cdot)$ is almost always specified as a quadratic or cubic function of per capita income. An obvious problem of working with parametric specifications such as those considered in this literature is the sensitivity of the results to the assumed functional form. There have been some recent attempts to allow $g(\cdot)$ to depend on x in a more flexible way. For example, Schmalensee et al. (1998) consider a piecewise linear specification with 10 segments, while Millimet et al. (2003) estimate (14) as a partially linear model. However, those studies that document the existence of a EKC for particular pollutants using nonparametric methods do not assign standard errors to their estimators of the turning point, so they cannot be used to create confidence intervals. In this section, the nonparametric methods previously described in the paper are used to estimate the location and size of the turning point of the EKC for NO_x, and provide standard errors for the estimators.[14]

[12] See, for instance, the World Bank's World Development Report 1992: Development and the Environment (IBRD, 1992). For some references regarding estimation of EKCs see footnote 1.

[13] Several reasons have been considered in the literature for the eventual decline in environmental degradation as income raises. Some of them are a negative income elasticity for pollution; increased levels of education, environmental awareness and openness of the political system; changes in the composition of consumption and production; better technologies, among others. See, for instance, Stern (1998) and Dasgupta et al. (2002). For a critical review of the EKC literature see Stern (1998) and Stern (2004).

[14] NO_x is a pollutant that receives considerably public policy attention and is one the most studied in this literature.

In this particular empirical application we follow the EKC literature and focus on estimation of the location and size of the turning point of the reduced-form model in (14) (e.g., List and Gallet,1999; Millimet et al., 2003), without claiming a causal interpretation to the relationship between per capita income and pollution. In addition, we also estimate (14) controlling for population density for the purpose of illustrating the results presented for partial means. A similar model controlling for population density using a quadratic form in income can be found in Selden and Song (1994). They argue that more densely populated areas are more likely to be concerned about reducing per capita emissions than areas where the population is more sparse.[15]

The data for emissions of NO_x and income used in this section is the same as the one analyzed in List and Gallet (1999) and Millimet et al. (2003). It comes originally from the US EPA in their National Air Pollutant Emission Trends, 1900-1994. It consists of data on emissions and per capita income for 48 US states from 1929 to 1994. One of the advantages of this data set is that it covers a long period of time, so it is more likely to cover both, the increasing and decreasing parts of the EKC.[16] Table 1 presents basic statistics of the variables used. Per capita emissions of NO_x are measured in thousands of short tons, per capita income in thousands of 1987 US dollars, and population density in habitants per square mile.

We first estimate a reduced form relation (i.e., without covariates), subsequently including population density in the model as a covariate. In both cases we estimate (14) as a partially linear model with the fixed effects as the linear part of the model. For comparison purposes we also present results for quadratic and cubic specifications of $g(\cdot)$ in (14), which are identical to those previously reported by List and Gallet (1999) and Millimet et al. (2003). The kernel estimators for the reduced-form case are based on a second-order Gaussian kernel.[17] The choice of bandwidth is always an issue when using nonparametric methods. In this application, the bandwidth is chosen as $\sigma = as_x n^{\delta-\eta}$, where $a = 1$, s_x is the sample standard deviation of x, δ helps to determine the order of the bandwidth, and $\eta > 0$ is a small number used for undersmoothing. This type of bandwidth has been previously used in the literature (e.g., Baltagi et al., 1996; Pagan and Ullah, 1999), and for our purposes it has the advantage that the order of the bandwidth can be specified directly. Below we analyze the sensitivity of the results to the choice of bandwidth by varying a. Given the kernel used, we base our estimator of the location of the turning point on a bandwidth of order $n^{-(1/7)-\eta}$, and our size estimator on a

[15] Other papers including population density as an additional explanatory variable include Panayotou (1993), Grossman and Krueger (1995), among others.

[16] For more information on the data see List and Gallet (1999) and Millimet et al. (2003).

[17] Although in the way theorem 1 was stated we would need to use higher order kernels, theorem 1 in Flores (2005) shows that it is possible to weaken those conditions when $k = 0$ to allow for second order kernels. We use this type of kernels in this reduced-form case because they are commonly used in this literature (e.g., Millimet et al, 2003).

bandwidth of order $n^{-(1/5)-\eta}$.[18,19] The asymptotic variances in Theorem 1 are estimated using "plug-in" estimators, so we substitute estimates for the unknown functions appearing there.

Figure 1 and Table 2 present results for the reduced-form case. Our location estimate of the turning point is 8,210 dollars, and the estimated level of emissions of NO_x at this point is 110.5 short tons. The estimated asymptotic standard errors for these estimators are 292.2 and 1.4, respectively. As a point of reference, in 1966 per capita income in Texas was 8,155 dollars.

We now let $g(\cdot)$ in (14) be a function of x and z, where z is population density. Theorem 1 requires the order of the kernel used to be greater than $\max\{2+k/2, k\}$, where k is the number of covariates used. Here we use a sixth order Gaussian kernel. Specifically, we use the product kernel $K(u,v) = \overline{K}(u)\,\overline{K}(v)$, with $\overline{K}(\zeta) = \frac{1}{8}\left(15 - 10\zeta^2 + \zeta^4\right)\phi(\zeta)$, and $\phi(\zeta)$ the standard normal density function. Based on our discussion of the bandwidth conditions in section 2, the order of the bandwidth used for estimation of the location of the turning point is $n^{-(1/15)-\eta}$, and the one used for estimation of the size is $n^{-(1/13)-\eta}$, where as before η is used to undersmooth. As before, the asymptotic variances in Theorem 1 are estimated using plug-in estimators. For this purpose, it is helpful to write the integral $\int \psi(\alpha_0, x)\,dx$ appearing in both variances as $E\left[f_0(\alpha_0|x)^{-1}\tau^2(x)\,Var\left[y|r=(\alpha_0,x)\right]\right]$, where $f_0(\alpha_0|x)$ is the conditional density of α given x, and the expectation is taken over the covariates. Thus, a plug-in estimator of this term is $n^{-1}\sum_{i=1}^{n}\{\widehat{f}(\widehat{\alpha}|x_i)^{-1}\tau^2(x_i)\,\widehat{Var}\left[y|r=(\widehat{\alpha},x_i)\right]\}$, where $\widehat{f}(\widehat{\alpha}|x_i)$ and $\widehat{Var}\left[y|r=(\widehat{\alpha},x_i)\right]$ are nonparametric estimators of the corresponding unknown functions.

Figure 2 and Table 3 present the results for this case. The estimated turning point based on our nonparametric estimator is 8,090 dollars, with a standard error of 310.7; and the estimated size of the turning point is 110.1, with a standard error of 1.2. These results are not very different from the ones obtained before in the reduced-form models. Selden and Song (1994), using parametric methods, also obtain that conditioning on population density does not affect their results considerably.

To check the sensitivity of the results to different choices of σ, we now vary a in the interval $[0.5, 2]$.[20] Table 4 presents the results. In general, the results are not drastically changed by the choice of bandwidth, except for the case when controlling for population density and $a = 2$. This may suggest that, not surprisingly, bandwidth choice is more important in the presence of covariates.

Finally, note that both estimates of the location of the turning point based on parametric models are above the ones from the nonparametric model, especially the ones from the quadratic

[18] The optimal bandwidth for a NW regression estimator used to estimate the d derivative of a regression function with q regressors and using a kernel of order s is of order $n^{-1/[2(d+s)+q]}$.

[19] In order to avoid boundary problems the nonparametric estimation is performed in the interval $[\min(x+\sigma), \max(x-\sigma)]$, where x is per capita income and σ is the bandwidth used. Other ways to proceed are the use of boundary kernels (e.g., Gasser and Müller, 1979) or local polynomial kernel estimators.

[20] Outside this range the nonparametric estimates of the EKC look either too undersmoothed or oversmoothed to be considered reasonable estimates of the curve.

specification commonly used in this literature. Even though in this particular application the estimates from the cubic specification are relatively close to the nonparametric ones, our approach allows the outcome variable to depend on the regressors on a more flexible way and are therefore less sensitive to functional form assumptions. This can be more relevant in other applications, as exemplified in the following section.

4 Monte Carlo

This section analyzes the finite properties of the estimators presented in this paper through a Monte Carlo study. First, we consider the case when the treatment level is assumed to be randomly assigned; and second, the case when we control for an additional covariate. Although this latter case has not been considered before in the literature, there are a few simulation results on estimation of the location and size of the maximum of a regression function in the statistics literature.[21] The simulations presented here for the first case differ from the ones in the statistics literature in various ways. First, we intend our simulation design to be closer to situations found in empirical research by basing our design on the same data set used in the previous section. Also, we consider larger sample sizes, a larger number of repetitions, and present a larger set of summary statistics of the simulation results including those regarding estimation of the asymptotic variance of our estimator, which has not been done before.

For simplicity, in this section we ignore the panel-data nature of the original data and pool all observations. We start with the case of randomly-assigned treatment doses. The functional forms considered for $g(t) = E[Y|T = t]$ are:

$$g_1(t) = 0.07 + 0.025\sin(0.5t) + 0.15e^{\left\{-0.15(t-8)^2\right\}} \quad (15a)$$
$$g_2(t) = 0.2 + 0.005\sin(0.75t - 5) - 0.001(t - 11)^2 \quad (15b)$$
$$g_3(t) = 0.09 + 0.05\sin(0.5t - 13) + 0.15e^{\left\{-0.02(4t-35)^2\right\}} \quad (15c)$$

The parameter values $(\alpha_0, \mu(\alpha_0))$ for each function are: $(7.7968, 0.2019)$, $(9.7418, 0.2021)$ and $(8.548, 0.2059)$, respectively. These parameter values were chosen to be close to the estimated turning point for the relation between emissions of NO_x and income obtained in section 3. Figures 3-5 show graphs for these functions. The first function has a sharp and symmetric peak. This function is similar to the one analyzed in Müller (1985, 1989). The second one has a smooth and asymmetric peak. Finally, the third function has also a sharp peak and is

[21] For example, Müller (1985) studies the performance of his estimators of location and size, which are based on the Gasser-Müller nonparametric estimator. He considers the fixed design case with equidistant points in the [0,1] interval and reports results for 100 repetitions with sample sizes 25 and 100. Müller (1989) uses 50 observations equidistantly in $[0, 1]$ and 200 repetitions focusing on the use by his estimators of a global versus a local bandwidth.

relatively highly nonlinear. While the case of a smooth peak represents a challenging case for our location estimator, the cases with a sharp peak are challenging for our size estimator.

We add a Gaussian error with standard deviation $s_\epsilon = 0.1$ to the functions in (15a)-(15c). For reference, the sample standard deviation of emissions of NO_x in our data is 0.07. For this case we consider five sample sizes: 100, 300, 500, 1000 and 3000. In order to have a better idea of the noise-to-signal ratios in our simulations, figures 3-5 also show representative simulated samples of size 500 for each of the models considered.

As in section 3, in our simulations we use a second order Gaussian kernel and choose the bandwidths equal to $\sigma_1 = s_t n^{-(1/7)-\eta}$ and $\sigma_2 = s_t n^{-(1/5)-\eta}$ for estimation of the location and size of the maximum, respectively; where s_t is the sample standard deviation of per-capita income at each simulated sample and η is a small quantity chosen to undersmooth.[22] At each replication, we estimate the asymptotic variance of our estimators in Theorem 1 using a plug-in estimator in the same way we did in our empirical application. Finally, for comparison purposes, we also present results for estimation of α_0 and $\mu(\alpha_0)$ based on a cubic model of per-capita income.[23]

Tables 5-7 present some of the results for the models in (15a)-(15c) based on 10,000 repetitions.[24] In general, the conclusions to draw from these simulations are: i) The nonparametric estimator of the location of the peak performs better for sharp than for smooth peaks. ii) The more non-linear the true regression function is the better is to use our nonparametric estimators of location and size, as compared to those based on a cubic specification, even for relatively small sample sizes (e.g., 100). iii) For the smooth peak in (15b), our location estimator needs a larger sample size to outperform the cubic-based model. iv) As discussed in section 2, the variance of the location estimator is higher with a smooth peak than with a sharp peak. v) Our size estimator performs better for smooth peaks than for sharp ones.[25] vi) For the three models

[22] See footnote 17.

[23] As in the previous section, we restrict the search for the maximum to the interval $[\min(t_i) + \sigma_1, \max(t_i) - \sigma_1]$ in order to avoid boundary problems; where t_i are the per-capita-income observations from a given simulated sample. In order to make a better comparison between the cubic and our nonparametric model, we also restrict the search for the maximum in the cubic model to the same interval.

[24] For the purposes of this section we define an estimated function to be "monotonic" when using our nonparametric approach if either $\widehat{g}_{\sigma_1}(\min(t_i) + \sigma_1) > \widehat{g}_{\sigma_1}(t)$ or $\widehat{g}_{\sigma_1}(\max(t_i) - \sigma_1) > \widehat{g}_{\sigma_1}(t)$ for all $t \in (\min(t_i) + \sigma_1, \max(t_i) - \sigma_1)$; where t_i are the per-capita-income observations from a given simulated sample. Note that this definition would classify as "monotonic" a case in which the estimated function has a local maximum but the function evaluated at one of the boundaries is larger. For consistency, in the cubic case we defined an estimated function to be monotonic if: i) the maximum of the estimated cubic function is outside $[\min(t_i) + \sigma_1, \max(t_i) - \sigma_1]$; or ii) the value of the estimated cubic function at either $\min(t_i) + \sigma_1$ or $\max(t_i) - \sigma_1$ is greater than at any other point in the interior of $[\min(t_i) + \sigma_1, \max(t_i) - \sigma_1]$.

[25] Given the relatively poor performance of the size estimators in the presence of sharp peaks, and following Müller (1989), Flores (2005) investigates the performance of our estimators when a local bandwidth is employed. For brevity, we have omitted those results. The use of local bandwidths improves the performance of the size estimator when the peak is sharp (as in Müller, 1989); however, it may negatively affect it in the case of smooth peaks. How much the performance of the size estimator improves with the use of a local bandwidth depends on how well the optimal local bandwidth is estimated.

and all sample sizes considered, the coverage rates of the confidence intervals for the location of the peak are higher than the nominal 95 and 90 percent. The coverage rates of the confidence intervals for the size are relatively low and decreasing as the sample size increases. However, in all cases they outperform those of the cubic model. vii) For the three models considered and for large sample sizes (e.g., 1000, 3000) the plug-in variance estimator of the location based on Theorem 1 tends to overestimate the standard error of the estimator. This happens for all sample sizes considered in cases with a sharp peak. On the other hand, the plug-in variance estimate of the size estimator provides a good approximation to its standard error. This may suggest that the plug-in estimator of the second derivative of the regression function evaluated at α_0 is not very accurate and is underestimating its true value.

We now consider controlling for an additional covariate: population density.[26] We consider two models, one having a sharp and symmetric peak and another one with a smooth and asymmetric peak. The true regression functions in this case are given by

$$\begin{aligned} g_1(t,x) &= -0.25 + 0.15e^{\left\{-0.15(t-9.5)^2\right\}} + 0.175e^{\left\{-0.025(t-0.1x)^2\right\}} \\ &\quad + 10000e^{\{-0.01x-10\}} \qquad (15d) \\ g_2(t,x) &= 0.01\sin(0.75t-5) - 0.002\,(t-0.01x-9)^2 + 10000e^{\{-0.01x-10\}} \qquad (15e) \end{aligned}$$

Following (4), the dose-response functions are given by $\mu_0(\bar{t}) = E\left[g(\bar{t},x)\right]$, where E is taken to be the empirical expectation of the population density variable based on the original data set. Using this approach, the true values of the parameters (α_0, γ_0) for the models based on (15d) and (15e) are $(9.2982, 0.2107)$ and $(9.4262, 0.2354)$, respectively. As before, we added a Gaussian error term with standard deviation $s_\varepsilon = 0.1$ to these models. The sample sizes analyzed are 100, 300, 500, 1000 and 2000. Figures 6 and 7 show the true dose-response function based on these functions along with a scatterplot of a representative simulated sample of size 500.

In order to satisfy our assumptions in Theorem 1 we base our estimates on the same sixth-order Gaussian kernel used in our empirical application. We choose the bandwidths using standardized data as $\sigma_1 = n^{-(1/15)-\eta}$ and $\sigma_2 = n^{-(1/13)-\eta}$ for estimation of the location and size of the maximum, respectively. As before, η is used to undersmooth. Finally, for comparison, we also estimate a cubic model in per-capita income controlling linearly for population density and evaluate it at the sample mean population density.

In the case of additional covariates one has to be especially careful regarding the use of nonparametric methods. Nonparametric estimators can become too noisy in regions where there is not enough data or there is a poor overlap between the treatment variable and the

[26]For referece, the correlation between population density and per capita income and emissions of NO_x is 0.3292 and -0.2275, respectively.

covariates. For instance, in our data we do not have observations with high (low) population density for values of per-capita income below 5000 (above 15,000) dollars (not shown in tables). One would expect nonparametric estimators to become noisier and more difficult to use in these regions, possibly ending up with a few very high estimates of the dose-response function which our estimator can erroneously identify as the maximum. For example, when in our simulations we allow the search for the maximum to be between $\min(t_i)$ and $\max(t_i)$, the standard deviation of the estimators is very large in some cases, and most importantly, we fail to find a maximum at all in many cases (i.e., the maximum is at the boundary). This illustrates the importance of having enough data and overlap between the treatment and covariates in order for the nonparametric estimators to work properly. To evaluate the performance of our estimators when we have "enough" data points and a "reasonable" overlap between our treatment and covariates, we simulate the models in (15d) and (15e) restricting the search for the maximum between the $25th$ and $75th$ sample percentiles of the treatment and trimming those observations with estimated joint density lower than 0.01.[27]

Table 8 and 9 present some of the results for the models in (15d) and (15e) based on 1000 repetitions. Some of the conclusions to draw are: i) As in the case of no additional covariates, our nonparametric estimator of location performs better for sharp than for smooth peaks. ii) For the sharp-peak case, the performance of our location and size estimators is better than that of the cubic model in terms of root MSE, median absolute error and coverage rates for all sample sizes considered. In some cases the differences in performance are very large. iii) For both models and all sample sizes considered, our size estimator performs better than the one based on the cubic model in terms of root MSE, median absolute error, bias and coverage rates. iv) For the smooth case in (15e), the location estimator based on the cubic model has a lower root MSE and median absolute error than our location estimator for all sample sizes considered. This may suggest that we need a large amount of data for our location estimator to perform better than the cubic-based one in this case. v) In both models, for the smaller (larger) sample sizes considered, our plug-in variance estimator for location tends to overestimate (underestimate) the standard error of the estimators. vi) It is important to have enough data and overlap of our treatment with the additional covariate(s) in order for our estimators to perform adequately; otherwise, we may be better off relying on parametric assumptions to extrapolate to those regions.

[27] The simulation results when we allow the search for the maximum to be between $\min(t_i)+\sigma_1$ and $\max(t_i)-\sigma_1$ are not shown in this paper. They are available in Flores (2005). The qualitative conclusions from these simulations are very similar to the ones in Tables (6) and (7), conditional on being able to find an interior maximum. However, the number of times the maximum is at the boundary in those simulations is substantially higher. For example, for a sample of size 3000 the maximum was at the boundary in about 50% (23%) of the cases with a sharp (smooth) peak.

5 Conclusions

This paper considers the continuous-treatment case and proposes nonparametric estimators for three objects of interest: the average dose-response function, and the location and size of the optimal dose. To identify these parameters we assume that units are assigned to different doses of the treatment based on an observed set of covariates and on unobserved components not correlated with the potential outcomes. Under this assumption the average dose-response function can be written as a partial mean (Newey, 1994). The proposed estimators are based on kernel estimators of partial means. This paper shows that the estimators for the location and size of the optimal dose are jointly asymptotically normal and uncorrelated. The asymptotic normality result can also be used in the case in which doses are randomly assigned and we want to find the location and size of the maximum of a regression function. Whether one needs to control for covariates or not, the scaling factors used for asymptotic normality of the estimators of the location and size of the optimal dose remain the same.

To illustrate the use of the tools developed in this paper, we estimate the location and size of the turning point of the environmental Kuznets curve for emissions of NO_x. We also carry out a Monte Carlo study partly based on the same data. The results show that the location and size estimators work well in practice, especially when compared to those from a parametric specification.

An important extension of the results presented in this paper deals with the case of selection into different treatment doses based on unobservables and the availability of a continuous instrument. In this case, one may use a control function approach similar to the one in Newey et al. (1999) to estimate the objects of interest analyzed in this paper. Another useful extension considers a more general class of dose-response functions, such as quantile dose-response functions.

6 Appendix

The proof of theorem 1 is based on the general framework developed in Newey (1994) (hereafter N) and uses some of his results from section 5. As in N, for a matrix B let $\|B\| = [tr\,(B'B)]^{1/2}$. Let $\Re$ be the compact set from assumption 6 where $h_{10}\,(r)$ is bounded away from zero; and let $\|h\,(r)\|_j = \max_{\ell \leq j} \sup_{r\in\Re} \left|\left|\partial^\ell h(r)/\partial r^\ell\right|\right|$ be the Sobolev norm used in N. Finally, let $C > 0$ be a generic constant which may take different values through the appendix.

Before proving Theorem 1, we prove the following Lemma.

Lemma A.1. Suppose that (i) Assumptions 1-6 are satisfied; (ii) for $\sigma_1 = \sigma_1\,(n)$, $\sigma_1 \to 0$; $n\sigma_1^{k+5}/\ln\,(n) \to \infty$; $n\sigma_1^{2k+1}/\left[\ln\,(n)\right]^2 \to \infty$ *and* $n\sigma_1^{2s+3} \to 0$. *Then,*

$$\sqrt{n\sigma_1^3}(\widehat{\alpha} - \alpha_0) \xrightarrow{d} \mathcal{N}\,(0, V_1)$$

with V_1 as given in Theorem 1.

Proof: From equation (12) in the text we have

$$\sqrt{n\sigma_1^3}(\widehat{\alpha}-\alpha_0) = -\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m_1(z_i,\alpha^*,\widehat{h}_{\sigma_1})}{\partial\alpha}\right]^{-1}\sqrt{n\sigma_1^3}\widehat{m}_{1,n}(\alpha_0) \tag{A.1}$$

for a mean value α^*, and where we let $\widehat{m}_{1,n}(\alpha_0) = \frac{1}{n}\sum_{i=1}^{n} m_1(z_i,\alpha_0,\widehat{h}_{\sigma_1})$. First we show $\frac{1}{n}\sum_{i=1}^{n}\partial m_1(z_i,\alpha^*,\widehat{h}_{\sigma_1})/\partial\alpha = \partial^2\widehat{\mu}_{\sigma_1}(\alpha^*)/\partial\alpha^2 \xrightarrow{p} \partial^2\mu_0(\alpha_0)/\partial\alpha^2$. Note that $\widehat{\alpha}$ is an extremum estimator that maximizes the objective function $\widehat{\mu}_{\sigma_1}(\overline{\alpha})$. We use Lemma 5.1 in N to show uniform convergence in probability of $\widehat{\mu}_{\sigma_1}(\overline{\alpha})$. Let $\overline{r} = (\overline{\alpha}, x)$ and $m_3(z,\alpha,h_0) = \tau(x)h_{20}(\overline{r})/h_{10}(\overline{r})$. Assumptions 3, 5 and 6(i) imply assumption (i) in Lemma 5.1. Conditions K, H and Y in Lemma 5.1 are satisfied by hypothesis. Choose $\varepsilon > 0$ small enough that $h_{10}(\overline{r})$ is bounded below for all $\overline{r}\in\Re$. Then, it is straightforward to show that for all $\alpha\in\mathcal{T}$ and $\|h-h_0\|\le\varepsilon$, $\|m_3(z,\alpha,h)-m_3(z,\alpha,h_0)\|\le C(\|h-h_0\|_0)^{\varepsilon}$. Finally, note that assumption (ii) in lemma A.1 implies $\ln(n)/n\sigma_1^{k+1}\to 0$, so all conditions in lemma 5.1 in N are satisfied. Therefore, $\sup_{\overline{\alpha}\in\mathcal{T}}\left|\widehat{\mu}_{\sigma_1}(\overline{\alpha})-\mu_0(\overline{\alpha})\right| \xrightarrow{p} 0$ and $\mu_0(\overline{\alpha})$ is continuous in $\mathcal{T}$, which along with assumption 5 imply that $\widehat{\alpha}\xrightarrow{p}\alpha_0$ (e.g., Theorem 2.1 in Newey and McFadden, 1994).

We use Lemma 5.1 in N again to show uniform convergence in probability of $\partial^2\widehat{\mu}_{\sigma_1}(\overline{\alpha})/\partial\alpha^2$. Let $m_4(z,\alpha,h_0) = \tau(x)\partial^2[h_{20}(\overline{r})/h_{10}(\overline{r})]/\partial\alpha^2$. As above, assumptions 2-5 and 6(i) imply assumption (i), K, H and Y in Lemma 5.1. Again, choose $\varepsilon > 0$ small enough that $h_{10}(\overline{r})$ is bounded below for all $\overline{r}\in\Re$. Using the quotient rule for derivatives along with mean value expansions (and assumptions 3 and 6(i)) one obtains that for all $\alpha\in\mathcal{T}$ and $\|h-h_0\|\le\varepsilon$, $\|m_3(z,\alpha,h)-m_3(z,\alpha,h_0)\|\le C(\|h-h_0\|_2)^{\varepsilon}$. Finally, the requirement that $\ln(n)/n\sigma_1^{k+5}\to 0$ in Lemma 5.1 is satisfied by hypothesis. Therefore, by Lemma 5.1 in N $\sup_{\overline{\alpha}\in\mathcal{T}}\left|\partial^2\widehat{\mu}_{\sigma_1}(\overline{\alpha})/\partial\alpha^2 - \partial^2\mu_0(\overline{\alpha})/\partial\alpha^2\right| \xrightarrow{p} 0$ and $\partial^2\mu_0(\overline{\alpha})/\partial\alpha^2$ is continuous in $\mathcal{T}$. Given that $\widehat{\alpha}\xrightarrow{p}\alpha_0$, we get $\partial^2\widehat{\mu}_{\sigma_1}(\alpha^*)/\partial\alpha^2 \xrightarrow{p} \partial^2\mu_0(\alpha_0)/\partial\alpha^2$, with $\partial^2\mu_0(\alpha_0)/\partial\alpha^2\neq 0$ by assumption 5.

We now show $\sqrt{n}\sigma_1^{3/2}\widehat{m}_{1,n}(\alpha_0)\xrightarrow{d}\mathcal{N}(0,M_1)$. To simplify notation, in what follows let $h^{(d)}(r)$ denote the d-th derivative of $h(r)$ with respect to α. Let $r_0 = (\alpha_0, x)$ and $r(\phi) = (\alpha_0,\phi)$. Define the following functions

$$\begin{aligned} & D(z,h,h_0) \\ = \; & \frac{\tau(x)}{h_{10}(r_0)}\left(-g_0(r_0)h_1^{(1)}(r_0) + h_2^{(2)}(r_0) + \left(\frac{2h_{10}^{(1)}(r_0)g_0(r_0) - h_{20}^{(1)}(r_0)}{h_{10}(r_0)}\right)h_1(r_0)\right. \\ & \left. -\frac{h_{10}^{(1)}(r_0)}{h_{10}(r_0)}h_2(r_0)\right) \end{aligned} \tag{A.2}$$

$$w_{11}(\phi) = \frac{\tau(\phi)\widetilde{f}_0(\phi)}{h_{10}(r(\phi))}\left[-g_0(r(\phi)), 1\right] \tag{A.3}$$

$$w_{12}(\phi) = \frac{\tau(\phi)\,\widetilde{f}_0(\phi)}{h_{10}(r(\phi))}\left[\left(\frac{2h_{10}^{(1)}(r(\phi))\,g_0(r(\phi)) - h_{20}^{(1)}(r(\phi))}{h_{10}(r(\phi))}\right), \frac{h_{10}^{(1)}(r(\phi))}{h_{10}(r(\phi))}\right] \tag{A.4}$$

$m_1(h) = \int D(z,h,h_0)\,dF(z)$; $m_{11}(h) = \int w_{11}(\phi)\,h^{(1)}(r(\phi))\,d\phi$ and $m_{12}(h) = \int w_{12}(\phi)$ $h(r(\phi))\,d\phi$; where $h^{(1)}(r) = [h_1^{(1)}(r) \;\; h_2^{(1)}(r)]'$.

Our proof proceeds by showing each of the following steps hold:

$$\begin{aligned}
\sqrt{n}\sigma_1^{3/2}\widehat{m}_{1,n}(\alpha_0) &= \sqrt{n}\sigma_1^{3/2}n^{-1}\textstyle\sum_{i=1}^n\left[m_1(z_i,\alpha_0,\widehat{h}_{\sigma_1}) - m_1(z_i,\alpha_0,h_0)\right] + o_p(1) && \text{(A.5.a)}\\
&= \sqrt{n}\sigma_1^{3/2}[m_1(\widehat{h}_{\sigma_1}) - m_1(h_0)] + o_p(1) && \text{(A.5.b)}\\
&= \sqrt{n}\sigma_1^{3/2}[m_{11}(\widehat{h}_{\sigma_1}) - m_{11}(h_0)] + o_p(1) \xrightarrow{d} \mathcal{N}(0, M_1) && \text{(A.5.c)}
\end{aligned}$$

Let $\psi_i = m_1(z_i,\alpha_0,h_0) = \tau(x_i)\,\partial g_0(\alpha_0,x_i)/\partial\alpha$, where $E(\psi_i) = \partial\mu_0(\alpha_0)/\partial\alpha = 0$ and $E(\psi_i^2) < \infty$ given assumptions 3 and 6(i). By Lindberg-Levy CLT we have $\left[E(\psi_i^2)\,n\right]^{-1/2}\sum_{i=1}^n$ $\psi_i \to N(0,1)$. Thus, $\sigma_1^{3/2}n^{-1/2}\sum_{i=1}^n m_1(z_i,\alpha_0,h_0) \xrightarrow{p} 0$ given $\sigma_1 \to 0$, and (A.5.a) follows.

The equality in (A.5.b) follows by checking the conditions of Lemma 5.4 in N. Let $D(z,h) = D(z,h,h_0)$ be as given in (A.2). $D(z,h)$ is linear in h on $\{h : \|h\|_1 < \infty\}$, so condition (i) in Lemma 5.4 is satisfied. As before, choose $\varepsilon > 0$ small enough that $h_{10}(r_0)$ is bounded below for all $r_0 \in \Re$. Then, by repeated used of the triangle inequality and the mean value theorem for functionals we have that for all h with $\|h - h_0\|_1 < \varepsilon$ (also note that, given our assumptions, $\|h - h_0\|_1 < \varepsilon$ implies $\|h\|_1 < \infty$ by $\|h_0\|_1 < \infty$)

$$\begin{aligned}
& ||m(z,\alpha_0,h) - m(z,\alpha_0,h_0) - D(z,h-h_0)|| \\
\leq\; & |\tau(x)|\left|\frac{h_2^{(1)}}{h_1h_{10}} - \frac{h_{20}^{(1)}}{h_{10}^2}\right||h_1 - h_{10}| + |\tau(x)|\left|\frac{1}{h_{10}^2}\right|\left|h_1^{(1)} - h_{10}^{(1)}\right||h_2 - h_{20}| \\
& + |\tau(x)|\left|\frac{h_1^{(1)}h_2(h_1 + h_{10})}{h_1^2h_{10}^2} - \frac{2h_{10}^{(1)}h_{20}}{h_{10}^3}\right||h_1 - h_{10}| \\
\leq\; & C\,\|h - h_0\|_1\,\|h - h_0\|_0
\end{aligned}$$

where before the last inequality all functions are evaluated at r_0. Hence, assumption (ii) in Lemma 5.4 is satisfied with $\Delta_1 = 1$, $\Delta_2 = 0$. Condition (iii) in Lemma 5.4 is also satisfied since by using Cauchy-Schwartz and triangle inequalities we have $|D(z,h)| \leq |C_1h_1^{(1)}(r_0) + C_2h_2^{(1)}(r_0)| + |C_3h_1(r_0) + C_4h_2(r_0)| \leq C\,\|h\|_1$. The rate hypothesis in Lemma 5.4 require that for $\eta_n^j = [\ln(n)/(n\sigma_1^{k+1+2j})]^{1/2} + \sigma^s$ we have a) $\eta_n^1 \to 0$; b) $\sqrt{n}\sigma_1^{3/2}\eta_n^1\eta_n^0 \to 0$; and c) $\sqrt{n}\sigma_1^{k+1/2} \to \infty$. $\sigma_1 \to 0$ and $n\sigma_1^{k+5}/\ln(n) \to \infty$ imply (a). Also, $n\sigma_1^{2k+1}/[\ln(n)]^2 \to \infty$ implies (c). Write $\sqrt{n}\sigma_1^{3/2}\eta_n^1\eta_n^0 = [\{\ln(n)\}^2/n\sigma_1^{2k+1}]^{1/2} + \left[n\sigma_1^{2s+3}\sigma_1^2\ln(n)/n\sigma_1^{k+5}\right]^{1/2} + \left[n\sigma_1^{2s+3}\sigma_1^4\ln(n)/n\sigma_1^{k+5}\right]^{1/2} + \left[n\sigma_1^{2s+3}\right]^{1/2}\sigma_1^s$. Hence, our assumptions imply $\sqrt{n}\sigma_1^{3/2}\eta_n^1\eta_n^0 \to 0$. Therefore, the conditions of Lemma 5.4 in N are satisfied, and for $m_1(h) = \int D(z,h)\,dF(z)$ the equality in (A.5.b) follows.

Since $m_1(h) = \int D(z,h)\,dF(z) = \int w_{11}(\phi)\,h^{(1)}(r(\phi))\,d\phi + \int w_{12}(\phi)\,h(r(\phi))\,d\phi = m_{11}(h)$ $+m_{12}(h)$, to show the equality in (A.5.c) it suffices to show $\sqrt{n}\sigma_1^{3/2}[m_{12}(\widehat{h}_{\sigma_1}) - m_{12}(h_0)] \xrightarrow{p} 0$. Write $\sqrt{n}\sigma_1^{3/2}[m_{12}(\widehat{h}_{\sigma_1}) - m_{12}(h_0)] = \sqrt{n}\sigma_1^{3/2}[m_{12}(\widehat{h}_{\sigma_1}) - E\{m_{12}(\widehat{h}_{\sigma_1})\}] + \sqrt{n}\sigma_1^{3/2}[E\{m_{12}(\widehat{h}_{\sigma_1})\} - m_{12}(h_0)]$. Note that $w_{12}(\phi)$ is bounded, continuous a.e. and zero outside the compact set $\Re$ by assumptions 3 and 6. Cauchy-Schwartz inequality implies that $|m_{12}(h)| \leq C\,\|h\|_0$. Hence, by Lemma B.4 in N $E\{m_{12}(\widehat{h}_{\sigma_1})\} = m_{12}(E\{\widehat{h}_{\sigma_1}\})$; and therefore, $\sqrt{n}\sigma_1^{3/2}[E\{m_{12}(\widehat{h}_{\sigma_1})\} - m_{12}(h_0)] = \sqrt{n}\sigma_1^{3/2}[m_{12}(E\{\widehat{h}_{\sigma_1}\}) - m_{12}(h_0)] = \sqrt{n}\sigma_1^{3/2}[m_{12}(E\{\widehat{h}_{\sigma_1}\} - h_0)] \leq \sqrt{n}\sigma_1^{3/2}C\|E\{\widehat{h}_{\sigma_1}\}$ $-h_0\|_0 = O(\left[n\sigma_1^{2s+3}\right]^{1/2}) \to 0$, where we used Lemma B.2 in N in the last equality. Now we show $A_n = \sqrt{n}\sigma_1^{3/2}[m_{12}(\widehat{h}_{\sigma_1}) - E\{m_{12}(\widehat{h}_{\sigma_1})\}] \xrightarrow{p} 0$. Since $E[A_n] = 0$, the result follows by showing $\lim_{n\to\infty} Var(A_n) = 0$. Similar to the proof of Lemma 5.3 in N, let $\rho_{12}(r) = \sigma_1^{-k-1}\int w_{12}(\phi)\,[I \otimes K((r(\phi) - r)/\sigma_1)]\,d\phi$, where I is a 2×2 identity matrix. Then, $m_{12}(\widehat{h}_{\sigma_1}) = n^{-1}\sum_{i=1}^{n}\rho_{12}(r_i)\,q_i$ and $Var(A_n) = \sigma_1^3 Var(\rho_{12}(r_i)\,q_i) = \sigma_1^3 E[\rho_{12}(r_i)\,q_i q_i' \rho_{12}'(r_i)] + o(1)$, since given i.i.d. data and following similar steps as above we get $|E[\rho_{12}(r_i)\,q_i] - m_{12}(h_0)| = |E\{m_{12}(\widehat{h}_{\sigma_1})\} - m_{12}(h_0)| \leq O(\sigma_1^s)$, which implies $\sigma_1^3 E[\rho_{12}(r_i)\,q_i] \to 0$. By a change of variables $u_2 = (\phi - x_i)/\sigma_1$ we can write $\rho_{12}(r_i)$ as $\rho_{12}(r_i) = \sigma_1^{-1}\int w_{12}(x_i + \sigma_1 u_2)\,[I \otimes K((\alpha_0 - t_i)/\sigma_1,$ $u_2)]du_2$. Let $\Omega(r_i) = E[q_i q_i' | r_i]$. Given the properties of $w_{12}(\phi)$ and our assumptions on the kernel we obtain:

$$\begin{aligned}
& \sigma_1 E[\rho_{12}(r_i)\,q_i q_i' \rho_{12}'(r_i)] \\
= \; & \sigma_1 E[\rho_{12}(r_i)\,\Omega(r_i)\,\rho_{12}'(r_i)] \\
= \; & \sigma_1^2 \int\int \rho_{12}(\alpha_0 - \sigma_1 u_1, x)\,\Omega(\alpha_0 - \sigma_1 u_1, x)\,\rho_{12}'(\alpha_0 - \sigma_1 u_1, x)\,f_0(\alpha_0 - \sigma_1 u_1, x)\,du_1 dx \\
\to \; & \left[\int \{\widetilde{K}(u_1)\}^2 du_1\right]\left[\int w_{12}(x)\,\Omega(\alpha_0, x)\,w_{12}'(x)\,f_0(\alpha_0, x)\,dx\right] < \infty
\end{aligned}$$

where in the third equality we used the change of variable $u_1 = (\alpha_0 - t_i)/\sigma_1$, and in the fourth one the bounded convergence theorem. Given $\sigma_1 \to 0$, this implies $\sigma_1^3 E[\rho_{12}(r_i)\,q_i q_i' \rho_{12}'(r_i)] \to 0$. Therefore, $Var(A_n) \to 0$, so that $A_n \xrightarrow{p} 0$ and the equality in (A.5.c) follows.

Finally, we use lemma 5.3 in N to show the asymptotic normality result in (A.5.c). By definition $m_{11}(h) = \int w_{11}(\phi)\,h^{(1)}(r(\phi))\,d\phi$, where $w_{11}(\phi)$ is bounded, continuous a.e. and zero outside the compact set $\Re$ by assumptions 3 and 6. The rate conditions in Lemma 5.3 are $(n\sigma_1)^{1/2} \to \infty$ and $\left(n\sigma_1^{2s+3}\right)^{1/2} \to 0$, which are implied by our assumptions. Also, assumptions 2-4 and 6 directly satisfy the rest of the conditions in Lemma 5.3. Then, the asymptotic normality in (A.5.c) follows with M_1 given by $M_1 = \int w_{11}(x)\left[\Omega(\alpha_0, x) \otimes \left\{\int\{\widetilde{K}_{u_1}(u_1)\}^2 du_1\right\}\right] w_{11}'(x)$ $f_0(\alpha_0, x)dx = \left[\int\{\widetilde{K}_{u_1}(u_1)\}^2 du_1\right]\int \psi(\alpha_0, x)\,dx$, with $\psi(\alpha_0, x)$ as in Theorem 1.

Given $\partial^2\widehat{\mu}_{\sigma_1}(\alpha^*)/\partial\alpha^2 \xrightarrow{p} \partial^2\mu_0(\alpha_0)/\partial\alpha^2 \neq 0$ and $\sqrt{n}\sigma_1^{3/2}\widehat{m}_{1,n}(\alpha_0) \xrightarrow{d} \mathcal{N}(0, M_1)$, the conclusion follows by Slutsky's theorem. Q.E.D.

Proof of Theorem 1. From equation (12) in the text we first show $J_n(z_i, \beta^*, \widehat{h}_{\sigma_1}, \widehat{h}_{\sigma_2}, \sigma_1, \sigma_2)$

$\xrightarrow{P} J(\beta_0, h_0)$. The elements in the second column are $\partial m_1(z,\beta,\widehat{h}_{\sigma_1})/\partial\gamma = 0$ and $\partial m_2(z,\beta,\widehat{h}_{\sigma_2})/\partial\gamma = -1$. The first diagonal term equals $\partial^2\widehat{\mu}_{\sigma_1}(\alpha^*)/\partial\alpha^2$, which in the proof of Lemma A.1 we showed $\partial^2\widehat{\mu}_{\sigma_1}(\alpha^*)/\partial\alpha^2 \xrightarrow{p} \partial^2\mu_0(\alpha_0)/\partial\alpha^2$. The last term equals $(\sigma_2/\sigma_1^3)^{1/2}\partial\widehat{\mu}_{\sigma_2}(\alpha^*)/\partial\alpha$. We now show $(\sigma_2/\sigma_1^3)^{1/2}\partial\widehat{\mu}_{\sigma_2}(\alpha^*)/\partial\alpha \xrightarrow{p} 0$. For suitable mean value α^{**} between α^* and α_0 write

$$\left(\frac{\sigma_2}{\sigma_1^3}\right)^{1/2}\frac{\partial\widehat{\mu}_{\sigma_2}(\alpha^*)}{\partial\alpha} = \frac{1}{\sqrt{n\sigma_2^2\sigma_1^3}}\sqrt{n\sigma_2^3}\frac{\partial\widehat{\mu}_{\sigma_2}(\alpha_0)}{\partial\alpha} + \frac{\sqrt{\sigma_2}}{\sqrt{n\sigma_1^6}}\frac{\partial^2\widehat{\mu}_{\sigma_2}(\alpha^{**})}{\partial\alpha^2}\sqrt{n\sigma_1^3}(\alpha^*-\alpha_0) \quad \text{(A.6)}$$

Note that $\sqrt{n\sigma_2^3}\partial\widehat{\mu}_{\sigma_2}(\alpha_0)/\partial\alpha = \sqrt{n\sigma_2^3}\frac{1}{n}\sum_{i=1}^n m_1(z_i,\alpha_0,\widehat{h}_{\sigma_2}) = \sqrt{n\sigma_2^3}\widehat{m}_{1,n,\sigma_2}(\alpha_0)$. Since σ_2 satisfies all the assumptions imposed on σ_1, by (A.5a)-(A.5c) we have $\sqrt{n\sigma_2^3}\widehat{m}_{1,n,\sigma_2}(\alpha_0) = O_p(1)$. Also note that our assumptions on σ_1 and σ_2 imply $\left(n\sigma_2^2\sigma_1^3\right)^{1/2} = [\left(n\sigma_2^4\right)^{1/2}\left(n\sigma_1^6\right)^{1/2}]^{1/2} \to \infty$. Hence, the first term to the right of (A.6) is $o_p(1)$. Since $|\alpha^*-\alpha_0| \leq |\widehat{\alpha}-\alpha_0|$, it follows from lemma A.1 that $(n\sigma_1^3)^{1/2}(\alpha^*-\widehat{\alpha}) = O_p(1)$. In the proof of lemma A.1 we showed uniform convergence in probability of $\partial^2\widehat{\mu}_{\sigma_1}(\alpha)/\partial\alpha^2$. Following similar steps as there, and given σ_2 satisfies all the assumptions imposed on σ_1, we have $\sup_{\alpha\in\mathcal{T}}\left|\partial^2\widehat{\mu}_{\sigma_2}(\alpha)/\partial\alpha^2 - \partial^2\mu_0(\alpha)/\partial\alpha^2\right| \xrightarrow{p} 0$ and $\partial^2\mu_0(\alpha)/\partial\alpha^2$ is continuous in $\mathcal{T}$. Also, $\alpha^{**} \xrightarrow{p} \alpha_0$ follows from $\widehat{\alpha} \xrightarrow{p} \alpha_0$. Hence, $\partial^2\widehat{\mu}_{\sigma_2}(\alpha^*)/\partial\alpha^2 \xrightarrow{p} \partial^2\mu_0(\alpha_0)/\partial\alpha^2$. Given our assumptions on σ_1 and σ_2 imply $\left(\sigma_2[n\sigma_1^6]^{-1}\right)^{1/2} \to 0$, we obtain $(\sigma_2/\sigma_1^3)^{1/2}\partial\widehat{\mu}_{\sigma_2}(\alpha^*)/\partial\alpha \xrightarrow{p} 0$. Therefore, $J_n(\cdot) \xrightarrow{P} J$, where J is a diagonal matrix with elements $\partial^2\mu_0(\alpha_0)/\partial\alpha^2 \neq 0$ (by assumption 5) and -1.

We now show $\sqrt{n}D_n[\widehat{m}_n(z_i,\beta_0,\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2})] \xrightarrow{d} \mathcal{N}(0,M)$. In (A.5a)-(A.5c) we showed $\sqrt{n}\sigma_1^{3/2}\frac{1}{n}\sum_{i=1}^n m_1(z_i,\alpha_0,\widehat{h}_{\sigma_1}) = \sqrt{n}\sigma_1^{3/2}[m_{11}(\widehat{h}_{\sigma_1}) - m_{11}(h_0)] + o_p(1)$, where $m_{11}(h) = \int w_{11}(\phi) h^{(1)}(r(\phi))\,d\phi$ with $w_{11}(\phi)$ given in (A.3) and $h^{(1)}(r) = [h_1^{(1)}(r) \quad h_2^{(1)}(r)]'$. Similarly, in his proof of Theorem 4.1 N shows $\sqrt{n}\sigma_2^{1/2}\frac{1}{n}\sum_{i=1}^n m_2(z_i,\alpha_0,\widehat{h}_{\sigma_2}) = \sqrt{n}\sigma_2^{1/2}[m_2(\widehat{h}_{\sigma_2}) - m_2(h_0)] + o_p(1)$, where $m_2(h) = \int w_2(\phi) h(r(\phi))\,d\phi$ and $w_2(\phi) = w_{11}(\phi)$. For simplicity let $w(\phi) = w_{11}(\phi)$, which is bounded, continuous a.e. and zero outside the compact set $\Re$ by assumptions 3 and 6. Cauchy-Schwartz inequality implies that $|m_1(h)| \leq C\|h\|_1$. Hence, by Lemma B.4 in N $E\{m_{11}(\widehat{h}_{\sigma_1})\} = m_{11}(E\{\widehat{h}_{\sigma_1}\})$; and therefore, $\sqrt{n}\sigma_1^{3/2}[E\{m_{11}(\widehat{h}_{\sigma_1})\} - m_{11}(h_0)] = \sqrt{n}\sigma_1^{3/2}[m_{11}(E\{\widehat{h}_{\sigma_1}\}) - m_{11}(h_0)] = \sqrt{n}\sigma_1^{3/2}[m_{11}(E\{\widehat{h}_{\sigma_1}\} - h_0)] \leq \sqrt{n}\sigma_1^{3/2}C||E\{\widehat{h}_{\sigma_1}\} - h_0||_1 = O(\left[n\sigma_1^{2s+3}\right]^{1/2}) \to 0$, where we used Lemma B.2 in N in the last equality. A similar argument shows $\sqrt{n}\sigma_2^{1/2}[E\{m_2(\widehat{h}_{\sigma_2})\} - m_2(h_0)] \to 0$ since by assumption $n\sigma_2^{2s+1} \to 0$. Define $m^*(\widetilde{h},\overline{h}) = [m_{11}(\widetilde{h}), m_2(\overline{h})]'$. Then, our goal is to show $\sqrt{n}D_n[m^*(\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2}) - E\{m^*(\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2})\}] \xrightarrow{d} \mathcal{N}(0,M)$. We follow steps similar to those in the proof of Lemma 5.3 in N. Let $\rho^*(r_i) = [\rho_{\sigma_1}(r_i), \rho_{\sigma_2}(r_i)]'$, with $\rho_{\sigma_1}(r) = \sigma_1^{-k-2}\int w(\phi)[I\otimes K_{u_1}((r(\phi)-r)/\sigma_1)]\,d\phi$ and $\rho_{\sigma_2}(r) = \sigma_2^{-k-1}\int w(\phi)[I\otimes K((r(\phi)-r)/\sigma_2)]\,d\phi$, where I is a 2×2 identity matrix. Note that by a change of variables $u_2 = (\phi - x_i)/\sigma_1$ and $u_2 = (\phi - x_i)/\sigma_2$ we can write, respectively, $\rho_{\sigma_1}(r_i) = \sigma_1^{-2}\int w(x_i+\sigma_1 u_2)[I\otimes K_{u_1}((\alpha_0 - t_i)/\sigma_1, u_2)]du_2$ and $\rho_{\sigma_2}(r_i) = \sigma_2^{-1}\int w(x_i+\sigma_2 u_2)[I\otimes K((\alpha_0 - t_i)/\sigma_2, u_2)]du_2$. By Liapunov's CLT and given $m^*(\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2}) = n^{-1}\sum_{i=1}^n \rho^*(r_i)\,q_i$, to obtain asymptotic normality of $m^*(\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2})$ it is sufficient to show that (i) $n^{-1}E[\|D_n\rho^*(r_i)\,q_i\|^4]$

$\rightarrow 0$ and $Var(D_n\rho^*(r_i)\,q_i) \rightarrow M$. Using the C_r-inequality we can write $n^{-1}E[\|D_n\rho^*(r_i)\,q_i\|^4] \leq 2\{n^{-1}E[||\sigma_1^{3/2}\rho_{\sigma_1}(r_i)\,q_i||^4] + n^{-1}E[||\sigma_2^{1/2}\rho_{\sigma_2}(r_i)\,q_i||^4]\}$. Note that $n^{-1}E[||\sigma_1^{3/2}\rho_{\sigma_1}(r_i)\,q_i||^4 \leq \sigma_1^6 n^{-1}E[||\rho_{\sigma_1}(r_i)\,||^4 E[||q_i||^4|r_i] \leq \frac{C}{n}\sigma_1^6 E[||\rho_{\sigma_1}(r_i)\,||^4\{1+E[y^4|r_i]\}] = \frac{C}{n\sigma_1}\int\int||\int w(x+\sigma_1 u_2)$ $[I\otimes K_{u_1}(u_1,u_2)]\,du_2||^4\{1+E\left[y^4|r=(\alpha_0-u_1\sigma_1,x)\right]\}f_0(\alpha_0-u_1\sigma_1,x)\,du_1dx \leq C/n\sigma_1 \rightarrow 0$. By a similar argument we have that $n^{-1}E[||\sigma_2^{1/2}\rho_{\sigma_2}(r_i)\,q_i||^4] \leq C/n\sigma_2 \rightarrow 0$. Therefore, $n^{-1}E[\|D_n\rho^*(r_i)\,q_i\|^4] \rightarrow 0$.

By i.i.d. data, and following similar steps as above, note that $|E[\rho_{\sigma 1}(r_i)\,q_i] - m_{11}(h_0)\,| = |E\{m_{11}(\widehat{h}_{\sigma_1})\} - m_{11}(h_0)\,| \leq O(\sigma_1^s)$, which implies $\sigma_1^{3/2}E[\rho_{11}(r_i)\,q_i] \rightarrow 0$. By a similar argument we have $\sigma_2^{1/2}E[\rho_{\sigma_2}(r_i)\,q_i] \rightarrow 0$. Hence, we need to show $E[D_n\rho^*(r_i)\,q_iq_i'\rho^{*\prime}(r_i)\,D_n'] \rightarrow M$. Let $\Omega(r_i) = E[q_iq_i'|r_i]$. For the first diagonal term we obtain

$$
\begin{aligned}
& \sigma_1^3 E[\rho_{\sigma_1}(r_i)\,q_iq_i'\rho_{\sigma_1}'(r_i)] \\
= \;& \sigma_1^3 E[\rho_{\sigma_1}(r_i)\,\Omega(r_i)\,\rho_{\sigma_1}'(r_i)] \\
= \;& \sigma_1^4 \int\int \rho_{\sigma_1}(\alpha_0-\sigma_1 u_1,x)\,\Omega(\alpha_0-\sigma_1 u_1,x)\,\rho_{\sigma_1}'(\alpha_0-\sigma_1 u_1,x)\,f_0(\alpha_0-\sigma_1 u_1,x)\,du_1dx
\end{aligned}
$$

$$
\rightarrow \left[\int\{\widetilde{K}_{u_1}(u_1)\}^2du_1\right]\int\psi(\alpha_0,x)\,dx = M_1
$$

where in the second line we used the change of variable $u_1 = (\alpha_0 - t_i)\,/\sigma_1$ and in the third one the bounded convergence theorem, with $\psi(\alpha_0,x)$ as in Theorem 1. Following a similar approach we obtain for the second diagonal term $\sigma_2 E[\rho_{\sigma_2}(r_i)\,q_iq_i'\rho_{\sigma_2}'(r_i)] \rightarrow [\int\{\widetilde{K}(u_1)\}^2du_1]\int\psi(\alpha_0,x)$ $dx = M_2$. Finally, consider the covariance term. Let $a_n = (\sigma_2/\sigma_1)^{1/2}$, $b_n = (\sigma_1/\sigma_2)^{1/2}$ and $\Gamma(\alpha,x) = w(x)\,\Omega(\alpha,x)\,w'(x)\,f(\alpha,x)$, where $\Gamma(\alpha_0,x) = \psi(\alpha_0,x)$. Then we can write

$$
\begin{aligned}
& \sigma_1^{3/2}\sigma_2^{1/2}E[\rho_{\sigma_1}(r_i)\,q_iq_i'\rho_{\sigma_2}'(r_i)] \qquad\qquad (A.7)\\
= \;& \sigma_1^2\sigma_2\int\int\rho_{\sigma_1}(\alpha_0-\sqrt{\sigma_1\sigma_2}u_1,x)\,\Omega(\alpha_0-\sqrt{\sigma_1\sigma_2}u_1,x)\,\rho_{\sigma_2}'(\alpha_0-\sqrt{\sigma_1\sigma_2}u_1,x) \\
& f_0(\alpha_0-\sqrt{\sigma_1\sigma_2}u_1,x)\,du_1dx \\
= \;& \int\int\Gamma(\alpha_0-\sqrt{\sigma_1\sigma_2}u_1,x)\left[\widetilde{K}_{u_1}(a_nu_1)\,\widetilde{K}(b_nu_1)\right]du_1dx + o(1) \\
= \;& \left[\int\psi(\alpha_0,x)\,dx\right]\left[\int\widetilde{K}_{u_1}(a_nu_1)\,\widetilde{K}(b_nu_1)\,du_1\right] + \\
& \sqrt{\sigma_1\sigma_2}\left[\int\left\{\frac{\partial\Gamma(\alpha^*,x)}{\partial\alpha}\right\}dx\right]\left[\int u_1\widetilde{K}_{u_1}(a_nu_1)\,\widetilde{K}(b_nu_1)\,du_1\right] + o(1)
\end{aligned}
$$

for a mean value α^*. The first equality in (A.7) follows by a change of variable $u_1 = (\alpha_0 - t)\,/$ $\sqrt{\sigma_1\sigma_2}$. For small σ_1 and σ_2, the second equality comes from mean value expansions of $\sigma_1^2\rho_{\sigma_1}\left(\alpha_0-\sqrt{\sigma_1\sigma_2}u_1,x\right) = \int w(x+\sigma_1u_2)\,[I\otimes K_{u_1}(a_nu_1,u_2)]\,du_2$ and $\rho_{\sigma_2}\left(\alpha_0-\sqrt{\sigma_1\sigma_2}u_1,x\right)$ $= w(x+\sigma_2u_2)\,[I\otimes K(b_nu_1,u_2)]\,du_2$ around x. For small $\sqrt{\sigma_1\sigma_2}$, the last expression in (A.7) is obtained by a mean value expansion of $\Gamma\left(\alpha_0-\sqrt{\sigma_1\sigma_2}u_1,x\right)$ around α_0. Consider the second term in the last line of (A.7). This term is $o(1)$ since$\int\{\partial\Gamma(\alpha^*,x)\,/\partial\alpha\}\,dx \rightarrow \int\{\partial\psi(\alpha_0,x)\,/\partial\alpha\}$

$dx < \infty$ and our assumptions on the kernel imply $\sqrt{\sigma_1\sigma_2}\int u_1\widetilde{K}_{u_1}(a_n u_1)\widetilde{K}(b_n u_1)\,du_1 \to 0$. Regarding the leading term, by symmetry of $\widetilde{K}(u_1)$, $\widetilde{K}_{u_1}(u_1) = -\widetilde{K}_{u_1}(-u_1)$, which implies $\int \widetilde{K}_{u_1}(a_n u_1)\widetilde{K}(b_n u_1)\,du_1 = 0$. Thus, $\sigma_1^{3/2}\sigma_2^{1/2}E[\rho_{\sigma_1}(r_i)\,q_i q_i' \rho_{\sigma_2}'(r_i)] \to 0$. Then, using Liapunov CLT we obtain $\sqrt{n}D_n[\widehat{m}_n(z_i,\beta_0,\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2})] = \sqrt{n}D_n[m^*(\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2}) - E\{m^*(\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2})\}] + o_p(1) \xrightarrow{d} \mathcal{N}(0,M)$, where M is a diagonal matrix with elements M_1 and M_2.

Given $J_n^{-1}(\cdot) \xrightarrow{P} J^{-1}$ and $\sqrt{n}D_n[\widehat{m}_n(z_i,\beta_0,\widehat{h}_{\sigma_1},\widehat{h}_{\sigma_2})] \xrightarrow{d} \mathcal{N}(0,M)$, the conclusion follows by Slutsky's theorem. Q.E.D.

References

[1] Baltagi, B.; Hidalgo, J. and Li, Q. (1996). "A Nonparametric Test for Poolability using Panel Data", *Journal of Econometrics*, 75, 345-367.

[2] Chen, H.; Huang, M. and Huang, W. (1996). "Estimation of the Location of the Maximum of a Regression Function using Extreme Order Statistics", *Journal of Multivariate Analysis*, 57, 191-214.

[3] Cropper, M. and Griffiths, C. (1994). "The Interaction of Population Growth and Environmental Quality", *American Economic Review,* 84, 250-254.

[4] Dasgupta, S.; Laplante, B.; Wang, H. and Wheeler, D. (2002). "Confronting the Environmental Kuznets Curve", *Journal of Economic Perspectives,* 16, 147-168.

[5] Dawid, A. (1979) "Conditional Independence in Statistical Theory (with Discussion)" *Journal of the Royal Statistical Society, Series B*, 41, 1-31.

[6] Eddy, W. (1982). "The Asymptotic Distributions of Kernel Estimators of the Mode", *Z. Wahrsch. Verw. Gabiete*, 59, 279-290.

[7] Flores, C. A. (2005). "Estimation of Dose-Response Functions and Optimal Doses with a Continuous Treatment", Doctoral Dissertation, University of California at Berkeley.

[8] Flores-Lagunes, A.; Gonzalez, A.; and Neumann, T. (2007) "Estimation of the Effects of the Length of Exposure to a Training Program: The Case of Job Corps", IZA Discussion Paper No. 2846.

[9] Firpo, S. (2007). "Efficient Semiparametric Estimation of Quantile Treatment Effects", *Econometrica*, 75(1), 259-276.

[10] Fryges, H. (2006). "The Export-Growth Relationship: Estimating a Dose-Response Function", ZEW Discussion Paper No. 06-028.

[11] Fryges, H. and Wagner, J. (2007). "Exports and Productivity Growth: First Evidence from a Continuous Treatment Approach", IZA Discussion Paper No. 2782.

[12] Grossman, G. M. and Krueger, A. (1991). "Environmental Impacts of a North American Free Trade Agreement", *NBER Working Paper*, No. 3914.

[13] Grossman, G. M. and Krueger, A. (1995). "Economic Growth and the Environment", *Quarterly Journal of Economics*, 110(2), 353-377.

[14] Hausman, J. A. and Newey, W. K. (1995). "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss", *Econometrica*, 63(6), 1145-1476.

[15] Heckman, J.; LaLonde, R. and Smith, J. (1999). "The Economics and Econometrics of Active Labor Market Programs", in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 3A, Amsterdam: North-Holland, 1865-2097.

[16] Heckman, N.E. (1992). "Bump Hunting in Regression Analysis", *Statistics and Probability Letters*, 14, 141-152.

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

[18] Hirano, K.; Imbens, G. W. and Ridder, G. (2003). "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score", *Econometrica*, 71(4), 1161-1189.

[19] Imbens, G. W. (2000). "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, 87(3), 706-710.

[20] Imbens, G. W. (2004). "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review", *Review of Economics and Statistics,* 86(1), 4-29.

[21] Imbs, J. and Wacziarg, R. (2003). "Stages of Diversification", *American Economic Review*, 93(1), 63-86.

[22] Kluve, J.; Schneider, H.; Uhlendorff, A. and Zhao, Z. (2007). "Evaluating Continuous Training Programs using the Generalized Propensity Score", Working Paper, IZA.

[23] List, J.A. and Gallet, C.A. (1999). "The Environmental Kuznets Curve: Does One Size Fit All?", *Ecological Economics,* (31), 409-423.

[24] Millimet, D. L.; List, J. A. and Stengos, T. (2003). "The Environmental Kuztens Curve: Real Progress or Misspecified Models", *Review of Economics and Statistics*, 85(4), 1038-1047.

[25] Mitnik, O. (2007). "Intergenerational Transmission of Welfare Dependency: The Effects of Length of Exposure", Working Paper, University of Miami.

[26] Müller, H.-G. (1985). "Kernel Estimators of Zeros and Location and Size of Extrema of Regression Functions", *Scandinavian Journal of Statistics*, 12, 221-232.

[27] Müller, H.-G. (1989). "Adaptive Nonparametric Peak Estimation", *Annals of Statistics*, 17(3), 1053-1069.

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

[29] Newey, W. K. and McFadden (1994). "Large Sample Estimation and Hypothesis Testing", in R. F. Engle and D. L. McFadden, eds., *Handbook of Econometrics*, Vol. 4, North-Holland, 2111-2245.

[30] Newey, W. K.; Powell, J. L. and Vella, F. (1999). "Nonparametric Estimation of Triangular Simultaneous Equations Models", *Econometrica*, 67(3), 565-603.

[31] Neyman, J. (1923) "On the Application of Probability Theory to Agricultural Experiments: Essays on Principles" Translated in *Statistical Science*, 5, 465-80.

[32] Pagan, A. and Ullah A. (1999). *Nonparametric Econometrics*, Cambridge University Press.

[33] Panayotou, T. (1993). "Empirical Test and Policy Analysis of Environmental Degradation at Different Stages of Economic Development", *Working Paper,* WP238, Technology and Employment Programme, International Labour Office, Geneva.

[34] Parzen, E. (1962). "On Estimation of a Probability Density Function and Mode", *Annals of Mathematical Statistics*, 33(3), 1065-1076.

[35] Romano, J. P. (1988). "On Weak Convergence and Optimality of Kernel Density Estimates of the Mode", *Annals of Statistics*, 16(2), 629-647.

[36] Royer, H. (2003). "The Question Every Woman (and Some Men) Ponder: Does Maternal Age Affect Infant Health?", Unpublished manuscript, Department of Economics, University of California at Berkeley.

[37] Rubin, D. B. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66(5), 688-701.

[38] Rubin, D. B. (1978). "Bayesian Inference for Causal Effects: The Role of Randomization", *Annals of Statistics*, 6(1), 34-58.

[39] Schmalensee, R.; Stoker, T. M. and Judson, R. A. (1998). “World Carbon Dioxide Emissions: 1950-2050”, *Review of Economics and Statistics,* 80(1), 15-27.

[40] Selden T. M. and Song, D. (1994). “Environmental Quality and Development: Is there a Kuznets Curve for Air Pollution Emissions?”, *Journal of Environmental Economics and Management,* 27, 147-162.

[41] Stern, D. I. (1998). “Progress on the Environmental Kuznets Curve?”, *Environment and Development Economics,* 3, 173-196.

[42] Stern, D. I. (2004). “The Rise and Fall of the Environmental Kuznets Curve”, *World Development,* 32(8), 1419-1439.

[43] Ziegler, K. (2000). “Nonparametric Estimation of Location and Size of Maxima of Regression Functions in the Random Design Case based on the Nadaraya-Watson Estimator with Data-Dependent Bandwidths”, PhD Thesis, University of Munich, Germany.

Table 1. Basic Statistics. Number of observations: 3168.

	Per-capita		
	Income	Emissions of NO_X	Population density
Mean	9.10	0.0928	132.32
Std. Deviation	4.24	0.0735	198.74
Minimum	1.16	0.023	0.820
Maximum	22.46	1.136	1080.3
Some percentiles:			
5%	2.81	0.0347	4.72
25%	5.85	0.0514	25.41
50%	8.43	0.0759	58.75
75%	12.38	0.1066	131.86
95%	16.27	0.2064	647.83

Emissions in thousand of short tons; income in thousands of 1987 US dollars and population density in habitants per square mile.

Table 2. Estimated location and size of the turning point of the EKC for NO_x including state and year fixed effects but no additional covariates.

	Functional form for per capita income		
	Quadratic	Cubic	Nonparametric
Estimated Turning point	10.79 (0.8118)	8.66 (0.7223)	8.21 (0.2922)
Estimated level at turning point (at average fixed effects)	0.1105 (0.0032)	0.1111 (0.0014)	0.1105 (0.0014)

Standard errors in parenthesis. For location and size we use a Gaussian kernel. Based on theorem 1 the bandwidths used for estimation of location and size using standardized data are 0.2916 and 0.1840, respectively.

Table 3. Estimated location and size of the turning point of the EKC for NO_x including state and year fixed effects and controlling for Population Density.

	Functional form for per capita income		
	Quadratic	Cubic	Partial Mean
Estimated Turning point	11.9 (1.2349)	8.87 (1.0408)	8.09 (0.3107)
Estimated level at turning point (at average fixed effects)	0.1067 (0.0046)	0.1053 (0.0016)	0.1101 (0.0012)

Standard errors in parenthesis. For location and size we use a sixth-order Gaussian kernel. Based on theorem1 the bandwidths used for estimation of location and size using standardized data are 0.5794 and 0.5337, respectively. In the partial mean model per capita income and population density both enter nonparametrically.

Table 4. Sensitivity of results to bandwidth selected. In tables 2 and 3 the bandwidths are chosen *as* $\sigma = an^{-(1/\delta)-\eta}$, where δ satisfies the conditions in theorem 1, η is used to undersmooth and *a*=1. Here *a* is varied.

	No additional covariates		Controlling for Population Density	
a	Location	Size	Location	Size
2	7.56	0.1096	11.18	0.0923
1.75	7.54	0.1099	8.21	0.1065
1.5	7.6	0.1102	7.89	0.1085
1.25	7.86	0.1103	7.91	0.1114
1	*8.21*	*0.1105*	*8.09*	*0.1101*
0.75	8.85	0.1121	8.74	0.1121
0.5	9.19	0.1152	8.34	0.1157

Table 5. Simulation results for regression function g_1 with a sharp peak at 7.7968 and size 0.2019. Number of repetitions: 10,000.

Sample size	Number of monotonic fits[a]	Mean bias	Median bias	Root MSE	Median abs. error	St. dev. of estimators	Range of estimators	Mean Std. Error	Median Std. Error	Coverage Rate nom. 95%	Coverage Rate nom. 90%
					Nonparametric Estimator of Location						
100	76	-0.2882	-0.2644	0.6716	0.4087	0.6067	[4.0033, 16.8752]	1.0972	1.0022	0.9866	0.9733
300	6	-0.158	-0.1538	0.3287	0.2196	0.2882	[6.0702, 8.7516]	0.6298	0.615	0.9987	0.9945
500	2	-0.1178	-0.1136	0.2482	0.1657	0.2185	[6.744, 8.4803]	0.4946	0.4866	0.9995	0.9976
1000	0	-0.0725	-0.0717	0.1685	0.1151	0.1521	[7.0848, 8.2812]	0.3567	0.3539	0.9997	0.9988
3000	0	-0.0193	-0.0203	0.0937	0.063	0.0917	[7.4094, 8.1134]	0.2154	0.2149	1	0.9998
					Cubic Estimator of Location						
100	146	-0.8758	-0.8869	1.1483	0.9057	0.7427	[3.8122, 11.4044]	0.7543	0.6286	0.6995	0.5985
300	141	-0.8156	-0.8177	0.9075	0.8179	0.398	[4.5898, 8.791]	0.3892	0.3716	0.4221	0.3145
500	194	-0.8043	-0.8071	0.8589	0.8071	0.3014	[5.5749, 8.153]	0.2963	0.2885	0.2227	0.1504
1000	200	-0.7899	-0.7922	0.8173	0.7922	0.2102	[6.1521, 7.8928]	0.2068	0.2042	0.0409	0.0217
3000	84	-0.7853	-0.7852	0.7944	0.7852	0.1204	[6.4995, 7.4827]	0.1183	0.1178	0	0
					Nonparametric Estimator of Size						
100	76	-0.0269	-0.0268	0.0311	0.0268	0.0156	[0.1145, 0.235]	0.0146	0.0146	0.5441	0.4305
300	6	-0.0203	-0.0205	0.0225	0.0205	0.0096	[0.1467, 0.2204]	0.0093	0.0093	0.4104	0.298
500	2	-0.0177	-0.0176	0.0193	0.0176	0.0079	[0.1486, 0.2166]	0.0076	0.0076	0.3644	0.2519
1000	0	-0.0141	-0.014	0.0153	0.014	0.0059	[0.1641, 0.2094]	0.0057	0.0057	0.3119	0.2096
3000	0	-0.0097	-0.0098	0.0104	0.0098	0.0038	[0.1784, 0.2075]	0.0037	0.0037	0.2477	0.1616
					Cubic Estimator of Size						
100	146	-0.0339	-0.0339	0.0374	0.0339	0.0159	[0.1092, 0.2282]	0.0156	0.0156	0.4121	0.2999
300	141	-0.0367	-0.0368	0.0378	0.0368	0.009	[0.132, 0.2054]	0.0089	0.0089	0.0172	0.0083
500	194	-0.0372	-0.0372	0.0379	0.0372	0.007	[0.1391, 0.1923]	0.0069	0.0069	0.0005	0.0003
1000	200	-0.0376	-0.0376	0.0379	0.0376	0.0049	[0.1459, 0.1823]	0.0049	0.0049	0	0
3000	84	-0.0378	-0.0378	0.0379	0.0378	0.0028	[0.1537, 0.1755]	0.0028	0.0028	0	0

a. For an explanation of what we mean by "monotonic fit" see footnote 24 in text.

Table 6. Simulation results for regression function g_2 with a smooth peak at 9.7418 and size 0.2021. Number of repetitions: 10,000.

Sample size	Number of monotonic fits[a]	Mean bias	Median bias	Root MSE	Median abs. error	St. dev. of estimators	Range of estimators	Mean Std. Error	Median Std. Error	Coverage Rate nom. 95%	Coverage Rate nom. 90%
Nonparametric Estimator of Location											
100	983	1.4983	1.1561	2.811	1.5476	2.3785	[3.9957, 19.6356]	2.69	1.8667	0.8036	0.7643
300	251	1.3202	0.9101	2.3459	1.1028	1.9392	[6.0027, 19.994]	1.9085	1.6136	0.8507	0.8163
500	77	1.134	0.7776	2.011	0.928	1.6609	[6.7419, 19.1878]	1.6164	1.4557	0.8767	0.8462
1000	7	0.849	0.6197	1.5004	0.7129	1.2372	[7.2079, 19.2959]	1.3079	1.2347	0.9218	0.8923
3000	0	0.5184	0.4321	0.8718	0.4841	0.701	[8.4736, 14.8575]	0.9246	0.8901	0.9742	0.9451
Cubic Estimator of Location											
100	921	1.1186	1.1191	2.2035	1.5387	1.8986	[4.2016, 19.7164]	1.9042	1.5334	0.7945	0.7305
300	222	1.2033	1.1506	1.7083	1.1975	1.2127	[7.1026, 18.8452]	1.2282	1.1112	0.8012	0.7276
500	67	1.1861	1.1688	1.5205	1.1771	0.9514	[7.3492, 16.416]	0.9492	0.8975	0.7509	0.6582
1000	11	1.1781	1.1718	1.361	1.1718	0.6814	[8.6094, 13.665]	0.6767	0.6596	0.6021	0.4766
3000	0	1.1753	1.1668	1.2386	1.1668	0.391	[9.5347, 12.4526]	0.3954	0.3911	0.148	0.0859
Nonparametric Estimator of Size											
100	983	0.0054	0.0046	0.0156	0.01	0.0147	[0.1563, 0.2985]	0.0167	0.016	0.9589	0.9183
300	251	0.0022	0.0019	0.0096	0.0062	0.0093	[0.1711, 0.2949]	0.0109	0.0106	0.9723	0.9355
500	77	0.0014	0.0012	0.0077	0.0051	0.0076	[0.1799, 0.2373]	0.0088	0.0087	0.9694	0.9366
1000	7	0.0004	0.0003	0.0058	0.0039	0.0058	[0.182, 0.2253]	0.0067	0.0066	0.973	0.9405
3000	0	-0.00004	-0.0001	0.0039	0.0027	0.0039	[0.184, 0.2187]	0.0043	0.0043	0.9718	0.9329
Cubic Estimator of Size											
100	921	0.0034	0.003	0.0146	0.0096	0.0142	[0.1584, 0.2583]	0.0155	0.0151	0.961	0.9168
300	222	-0.0005	-0.0005	0.0082	0.0055	0.0081	[0.1701, 0.2386]	0.0085	0.0083	0.9592	0.9135
500	67	-0.0011	-0.0012	0.0065	0.0043	0.0064	[0.1793, 0.2317]	0.0065	0.0064	0.9501	0.899
1000	11	-0.0017	-0.0017	0.0048	0.0032	0.0045	[0.184, 0.219]	0.0045	0.0045	0.9354	0.8764
3000	0	-0.002	-0.002	0.0033	0.0023	0.0026	[0.1895, 0.2095]	0.0026	0.0026	0.8772	0.8003

a. For an explanation of what we mean by "monotonic fit" see footnote 24 in text.

Table 7. Simulation results for regression function g_3 with peak at 8.5480 and size 0.2059. Number of repetitions: 10,000.

Sample size	Number of monotonic fits[a]	Mean bias	Median bias	Root MSE	Median abs. error	St. dev. of estimators	Range of estimators	Mean Std. Error	Median Std. Error	Coverage Rate nom. 95%	Coverage Rate nom. 90%
					Nonparametric Estimator of Location						
100	1719	-0.9605	-0.8996	1.5344	0.9182	1.1967	[3.1597, 19.5984]	2.6788	1.4978	0.937	0.9048
300	590	-0.6578	-0.6149	0.811	0.6152	0.4744	[2.9875, 17.4201]	1.0967	0.9924	0.9868	0.9713
500	198	-0.5115	-0.4959	0.5844	0.4959	0.2828	[3.4566, 17.8204]	0.8216	0.7905	0.9963	0.9877
1000	13	-0.373	-0.3676	0.4041	0.3676	0.1557	[7.4337, 8.708]	0.58	0.5697	0.9995	0.9968
3000	0	-0.2188	-0.2175	0.2324	0.2175	0.0785	[8.0223, 8.6256]	0.3357	0.3341	1	0.9996
					Cubic Estimator of Location						
100	2027	-2.1707	-2.2853	2.5535	2.3137	1.345	[3.2992, 19.057]	1.4614	1.0434	0.4524	0.3586
300	2183	-2.4295	-2.3697	2.5786	2.3706	0.8643	[2.9406, 11.6721]	1.0033	0.7607	0.2239	0.1465
500	2650	-2.4422	-2.364	2.5434	2.364	0.7103	[2.7918, 8.679]	0.7708	0.6284	0.0759	0.0399
1000	3566	-2.3994	-2.3457	2.4505	2.3457	0.4977	[3.0267, 7.5747]	0.5092	0.4603	0.0016	0.0005
3000	4706	-2.3738	-2.3599	2.3884	2.3599	0.2635	[4.7437, 6.9549]	0.2736	0.266	0	0
					Nonparametric Estimator of Size						
100	1719	-0.0405	-0.0406	0.0431	0.0406	0.0146	[0.1094, 0.2218]	0.015	0.0148	0.2255	0.1505
300	590	-0.0366	-0.0367	0.0379	0.0367	0.01	[0.1304, 0.2091]	0.0096	0.0095	0.0414	0.0204
500	198	-0.0332	-0.0332	0.0342	0.0332	0.0084	[0.1452, 0.204]	0.0078	0.0078	0.0195	0.0091
1000	13	-0.028	-0.028	0.0287	0.028	0.0063	[0.1547, 0.2022]	0.006	0.006	0.0053	0.0025
3000	0	-0.0203	-0.0203	0.0207	0.0203	0.0041	[0.1699, 0.2006]	0.0039	0.0039	0.0011	0.0003
					Cubic Estimator of Size						
100	2027	-0.0463	-0.0468	0.0487	0.0468	0.0152	[0.1118, 0.2222]	0.0161	0.016	0.1751	0.1064
300	2183	-0.0517	-0.0518	0.0525	0.0518	0.0091	[0.1209, 0.1919]	0.0092	0.0091	0.0001	0.0001
500	2650	-0.0531	-0.0531	0.0536	0.0531	0.0071	[0.1242, 0.1835]	0.0071	0.007	0	0
1000	3566	-0.0542	-0.0542	0.0544	0.0542	0.005	[0.1313, 0.1712]	0.0049	0.0049	0	0
3000	4706	-0.0547	-0.0546	0.0547	0.0546	0.0028	[0.142, 0.1623]	0.0028	0.0028	0	0

a. For an explanation of what we mean by "monotonic fit" see footnote 24 in text.

Table 8. Simulation results for model based on regression function $g_1(t,x)$. In this case the dose-response function has a sharp peak at 9.2982 with size 0.2107. Number of repetitions: 1,000.

Sample size	Number of monotonic fits[a]	Mean bias	Median bias	Root MSE	Median abs. error	St. dev. of estimators	Range of estimators	Mean Std. Error	Median Std. Error	Coverage Rate nom. 95%	Coverage Rate nom. 90%
					Nonparametric Estimator of Location						
100	116	-0.5039	-0.2978	1.2923	0.635	1.1907	[5.20, 11.65]	2.7647	2.1197	1	0.9989
300	72	-0.2228	-0.0771	0.943	0.4366	0.9168	[5.78, 11.21]	1.2022	1.0034	0.9989	0.9806
500	55	-0.0975	0.0558	0.8542	0.3525	0.8491	[5.77, 11.23]	0.8072	0.705	0.9778	0.9503
1000	42	0.0655	0.0989	0.697	0.2763	0.6943	[6.02, 11.16]	0.5175	0.4457	0.9113	0.8622
2000	3	0.2148	0.2059	0.423	0.2628	0.3645	[6.39, 10.81]	0.2862	0.2831	0.8485	0.7904
					Cubic Estimator of Location						
100	104	-1.9593	-2.098	2.1765	2.098	0.9483	[4.46, 11.22]	0.9862	0.8015	0.3828	0.2991
300	23	-2.2953	-2.3391	2.3477	2.3391	0.4937	[5.66, 8.89]	0.4894	0.4535	0.0358	0.0225
500	7	-2.3244	-2.3385	2.3527	2.3385	0.3639	[5.81, 8.44]	0.3544	0.3356	0.003	0
1000	0	-2.3508	-2.355	2.3648	2.355	0.2571	[6.17, 7.98]	0.2434	0.2383	0	0
2000	0	-2.3487	-2.3522	2.3548	2.3522	0.17	[6.23, 7.57]	0.1692	0.1668	0	0
					Nonparametric Estimator of Size						
100	116	-0.0062	-0.0055	0.0292	0.0197	0.0285	[0.1236, 0.2966]	0.0205	0.0205	0.8167	0.75
300	72	-0.0115	-0.0118	0.0223	0.0153	0.0191	[0.1399, 0.2765]	0.0121	0.012	0.708	0.6164
500	55	-0.0127	-0.0125	0.0201	0.0143	0.0156	[0.1428, 0.2445]	0.0095	0.0095	0.6106	0.5407
1000	42	-0.0113	-0.0115	0.0165	0.0123	0.012	[0.161, 0.2448]	0.0069	0.0069	0.5449	0.4562
2000	3	-0.0091	-0.0093	0.0125	0.0096	0.0085	[0.174, 0.2272]	0.005	0.005	0.5045	0.4173
					Cubic Estimator of Size						
100	104	-0.1014	-0.0997	0.1065	0.0997	0.0325	[0.0134, 0.2113]	0.0198	0.0197	0.0201	0.0156
300	23	-0.1024	-0.103	0.104	0.103	0.0178	[0.0434, 0.1642]	0.0113	0.0113	0	0
500	7	-0.1023	-0.102	0.1033	0.102	0.0145	[0.0568, 0.1527]	0.0087	0.0087	0	0
1000	0	-0.1014	-0.1011	0.1019	0.1011	0.0101	[0.0767, 0.1383]	0.0062	0.0062	0	0
2000	0	-0.1022	-0.1023	0.1025	0.1023	0.0074	[0.0844, 0.1318]	0.0044	0.0044	0	0

a. For an explanation of what we mean by "monotonic fit" see footnote 24 in text.

Table 9. Simulation results for model based on regression function $g_2(t,x)$. In this case the dose-response function has a smooth peak at 9.4262 with size 0.2354. Number of repetitions: 1,000.

Sample size	Number of monotonic fits[a]	Mean bias	Median bias	Root MSE	Median abs. error	St. dev. of estimators	Range of estimators	Mean Std. Error	Median Std. Error	Coverage Rate nom. 95%	nom. 90%
Nonparametric Estimator of Location											
100	109	-0.253	-0.1978	1.7197	1.2668	1.7019	[4.67, 13.67]	5.4938	3.5125	0.9989	0.9989
300	99	-0.085	0.1026	1.4777	0.9019	1.476	[5.24, 12.83]	2.3707	2.0502	0.9945	0.98
500	77	0.1089	0.283	1.344	0.8122	1.3404	[5.89, 12.73]	1.7566	1.5791	0.9772	0.9523
1000	79	0.2315	0.3705	1.2182	0.6814	1.1966	[5.88, 12.36]	1.2328	1.1311	0.9446	0.911
2000	57	0.1954	0.4204	1.1654	0.6215	1.1495	[5.94, 11.94]	0.8518	0.7976	0.9109	0.8717
Cubic Estimator of Location											
100	95	-0.5062	-0.697	1.3661	1.0799	1.2695	[5.60, 12.50]	1.431	1.1519	0.7901	0.737
300	11	-0.4032	-0.5036	0.9522	0.7385	0.863	[7.24, 12.17]	0.841	0.7625	0.8079	0.7371
500	0	-0.4081	-0.4721	0.788	0.6129	0.6745	[7.32, 11.60]	0.6445	0.6048	0.784	0.717
1000	0	-0.4791	-0.5017	0.668	0.5353	0.4657	[7.74, 11.25]	0.4364	0.4233	0.701	0.62
2000	0	-0.4976	-0.5132	0.5921	0.5173	0.321	[8.01, 10.30]	0.3019	0.2959	0.568	0.471
Nonparametric Estimator of Size											
100	109	0.021	0.0213	0.0346	0.0247	0.0276	[0.1682, 0.3423]	0.0216	0.0214	0.7699	0.6723
300	99	0.014	0.0143	0.0223	0.0159	0.0174	[0.1892, 0.3064]	0.013	0.013	0.7314	0.6349
500	77	0.0113	0.0112	0.0184	0.0128	0.0146	[0.1995, 0.2951]	0.0102	0.0102	0.7075	0.6381
1000	79	0.0102	0.0101	0.0149	0.0111	0.0108	[0.2079, 0.2791]	0.0074	0.0074	0.6363	0.5581
2000	57	0.0091	0.0089	0.0124	0.0094	0.0085	[0.2165, 0.2726]	0.0053	0.0053	0.5493	0.474
Cubic Estimator of Size											
100	95	-0.0643	-0.0635	0.071	0.0635	0.0303	[0.077, 0.2714]	0.0185	0.0184	0.168	0.1249
300	11	-0.068	-0.0683	0.0702	0.0683	0.0176	[0.1053, 0.2218]	0.0103	0.0103	0.002	0.001
500	0	-0.0681	-0.0685	0.0694	0.0685	0.0132	[0.1243, 0.2106]	0.0079	0.0079	0	0
1000	0	-0.0681	-0.0681	0.0687	0.0681	0.0096	[0.139, 0.1956]	0.0056	0.0056	0	0
2000	0	-0.0682	-0.0683	0.0685	0.0683	0.0065	[0.148, 0.1875]	0.0039	0.0039	0	0

a. For an explanation of what we mean by "monotonic fit" see footnote 24 in text.

Figure 1. Curves used for estimation of the location and size of the turning point of the EKC for NOx, including state and year effects but no additional covariates

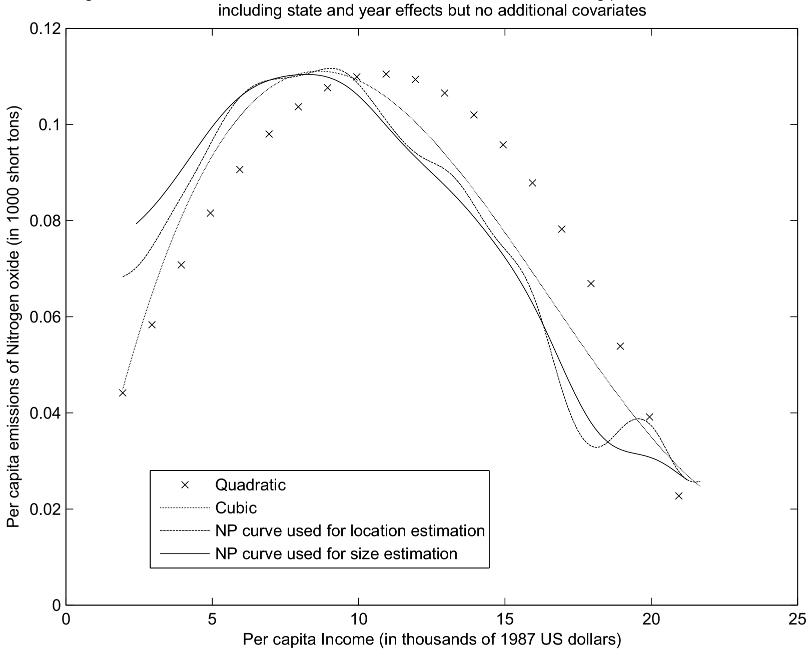

Figure 2. Curves used for estimation of the location and size of the turning point of the EKC for NOx, including state and year effects and controlling for population density

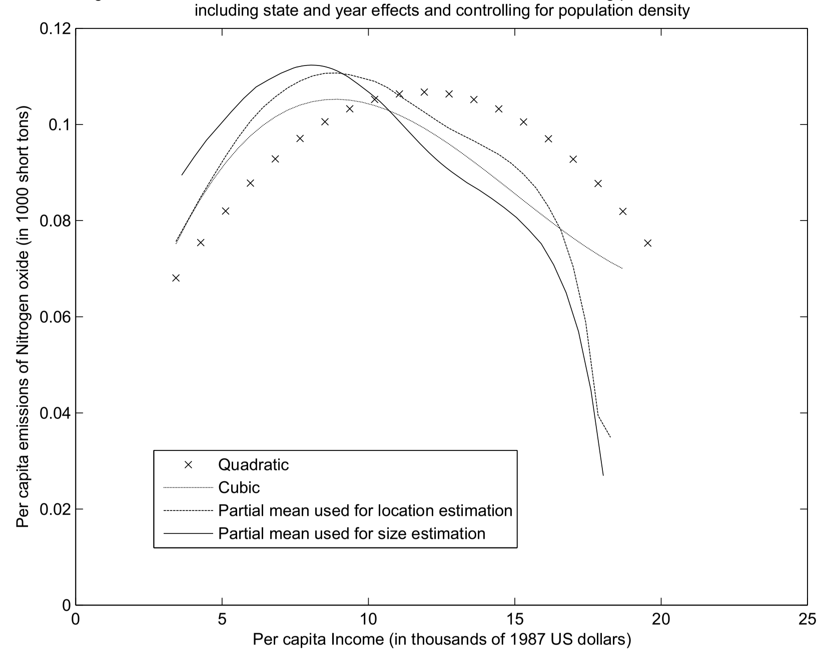

Figure 3. Regression curve g_1 and a representative simulated sample of size 500.

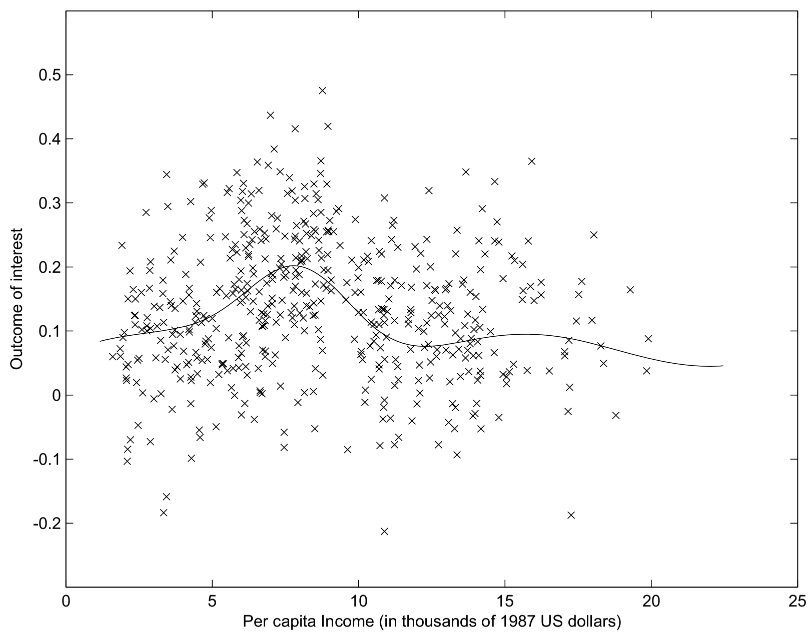

Figure 4. Regression curve g_2 and a representative simulated sample of size 500.

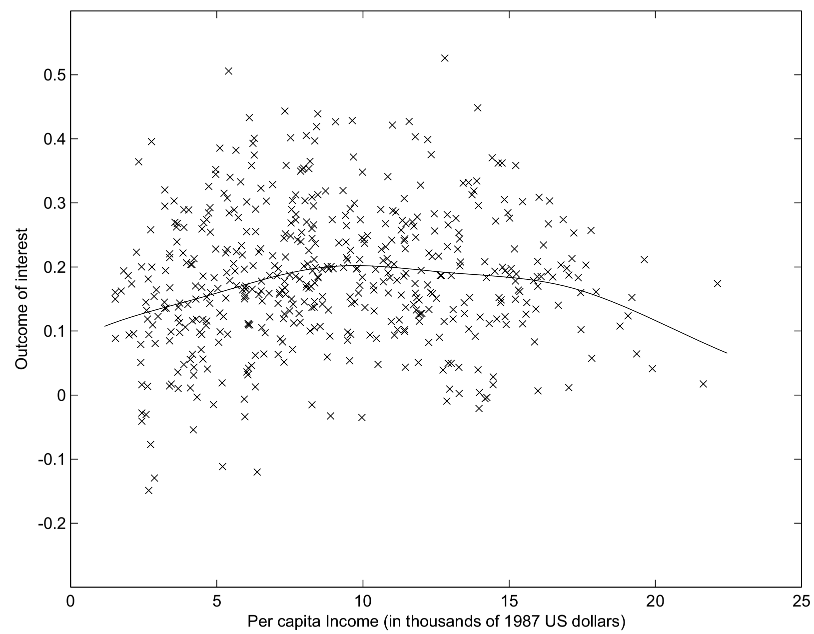

Figure 5. Regression curve g_3 and a representative simulated sample of size 500.

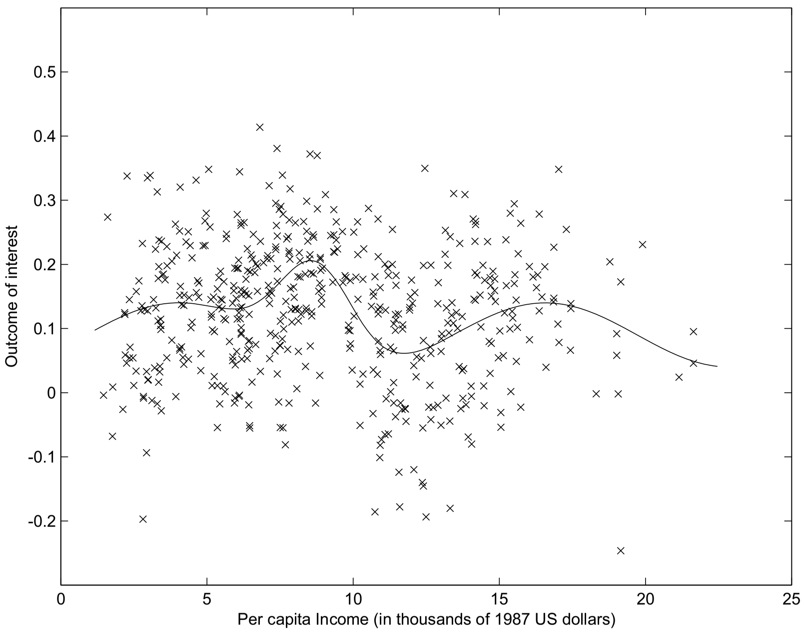

Figure 6. True dose-response function based on $g_1(t,x)$, along with a representative simulated sample of size 500.

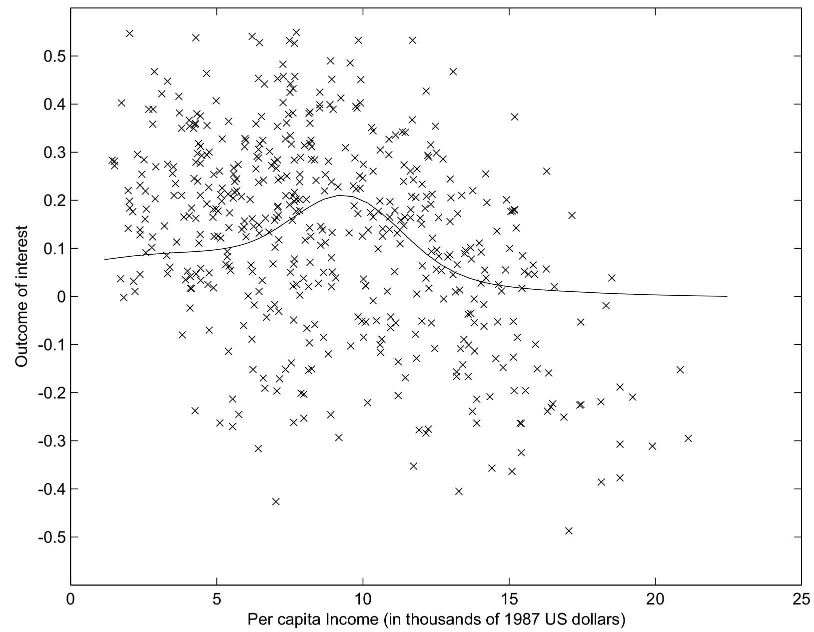

Figure 7. True dose-response function based on $g_2(t,x)$, along with a representative simulated sample of size 500.