Journal of the Korean Statistical Society (2001), **30**: *2, pp 265–280*

On Nonparametric Estimation of Data Edges †

Byeong U. Park[1]

Abstract

Estimation of the edge of a distribution has many important applications. It is related to classification, cluster analysis, neural network, and statistical image recovering. The problem also arises in measuring production efficiency in economic systems. Three most promising nonparametric estimators in the existing literature are introduced. Their statistical properties are provided, some of which are new. Themes of future study are also discussed.

Keywords: Boundary, frontier function, nonparametric function estimation, productivity analysis, Poisson process, Free Disposal Hull, Data Envelopment Analysis, local polynomial estimator, bandwidth.

1. Introduction

Let g be a real-valued smooth function defined on $\mathbb{R}^d$ and consider the set in $\mathbb{R}^{d+1}$:

$$\mathcal{S} = \{(x, y) \in \mathbb{R}^d \times \mathbb{R} \,:\, y \le g(x),\ x \in \mathcal{D}\}$$

for some $\mathcal{D} \subset \mathbb{R}^d$. We wish to estimate the boundary function g from a data set whose distribution is supported on only $\mathcal{S}$.

Estimation of a boundary is motivated by many practical problems. It may provide useful tools for determining classification schemes in statistical pattern recognition. It is closely related to the problem of estimating density level sets, which may be converted to that of clustering. The problem also arises in scatter-point image analysis. There, the function g represents, typically, the interface of areas of differing color tones, perhaps black above the boundary, where no data values are registered, and grey below. Korostelev and Tysbakov (1993) provide an excellent introduction to optimality issue in this type of problems.

†This research was supported by KOSEF through Statistical Research Center for Complex Systems.

[1]Department of Statistics,Seoul National University

Another type of applications of this problem may be found in the economic literature, especially on productivity analysis in the context of measuring efficiency of enterprises. In that context, the extremal value $g(x)$ represents a hypothetical upper limit to production or performance at a d-dimensional input level x, and the efficiency of a production unit with input x and output y is calculated by $g(x) - y$. The economic theory underlying efficiency analysis is based on the pioneering works of Koopmans (1951) and Debrue (1951) on activity analysis, and on the first empirical work of Farrell (1957). Shephard (1970) provides a modern economic formulation of the problem.

In parametric or semiparametric approaches some specific structural assumptions are imposed on the function g, such as $g(x) = \beta_0 + \beta^T x$, while in nonparametric approaches such restrictions are avoided and it is only assumed that g belongs to a function space of infinite dimension. We shall focus on the latter in this paper. Some important works with parametric or semiparametric treatments are the maximum likelihood estimation of Greene (1980), the instrumental variables methods of Hausman and Taylor (1981), and the semiparametric optimal procedures of Park and Simar (1994) and Park, Sickles and Simar (1998, 2000).

In this paper, three most promising nonparametric estimators are considered. They are the FDH(Free Disposal Hull; Deprins, Simar and Tulkens, 1984), the DEA(Data Envelopment Analysis; Farrell, 1957) and the local polynomial estimators(Hall, Park and Stern, 1998). The first two of these aim at boundary functions with special structures, namely monotonicity and/or concavity, which are very common in economic applications. The last one is for fairly general boundaries. These are introduced in the next section. The statistical properties which are now available on these estimators are summarized in Section 3 with some new results. A brief guide to the existing literature are provided there, too. Several problems for future study are discussed in Section 4.

2. The Nonparametric Estimators

2.1. The FDH estimator

Let $\mathcal{P} = \{(X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R} : i = 1, 2, \ldots\}$ be the data set. The FDH estimator of $\mathcal{S}$ was introduced by Deprins, Simar and Tulkens (1984). It relies on the *free disposability* assumption on $\mathcal{S}$, i.e. if $(x, y) \in \mathcal{S}$ then all pairs (x', y') such that $x' \geq x$ and $y' \leq y$ belong to $\mathcal{S}$, which is equivalent to the assumption that g is monotone nondecreasing. Here and below, the inequality $x' \geq x$ for d-dimensional vectors x and x' is understood componentwise. If $\mathcal{S}$ is interpreted as a production set, the totality of all combinations of inputs x and outputs y that are technically possible, then the free disposability of $\mathcal{S}$ means that, in that production process, if x can produce y then all $x' \geq x$ can produce all $y' \leq y$. The estimator of $\mathcal{S}$ is then defined as the free disposal hull of the set $\mathcal{X}$:

$$\hat{\mathcal{S}}_{\text{FDH}} = \bigcup_i FD\{(X_i, Y_i)\}$$

where $FD(z)$ for a point $z = (z_1, \ldots, z_{d+1})$ denotes the *free disposal set* of z:

$$FD(z) = \{(v_1, \ldots, v_{d+1}) \,:\, v_1 \geq z_1, \cdots, v_d \geq z_d,\ v_{d+1} \leq z_{d+1}\}.$$

The FDH estimator of $\mathcal{S}$ is the set under the "lowest" monotone step function that lies above all points (X_i, Y_i)'s. The lowest monotone step function is taken as the FDH estimator of g, i.e.

$$\hat{g}_{\text{FDH}}(x) = \max\{y \,:\, (x, y) \in \hat{\mathcal{S}}_{\text{FDH}}\}.$$

Note that this estimator may not be defined at x when there does not exist a point (X_i, Y_i) such that $X_i \leq x$. However, under fairly general data generating processes such as the one considered in Section 3, the probability that such events occur is negligible in the limit. Also, it is easy to see that, for a compact subset $\mathcal{K}$ in $\mathbb{R}^{d+1}$, the FDH estimator $\hat{\mathcal{S}}_{\text{FDH}}(\hat{g}_{\text{FDH}})$ is the nonparametric maximum likelihood estimator of $\mathcal{S}(g)$ on the class of all sets $\mathcal{S} \subset \mathcal{K}$ with monotone g, provided (X_i, Y_i) are iid with the uniform distribution on $\mathcal{S}$.

2.2. The DEA estimator

The DEA(Data Envelopment Analysis) approach was introduced by Farrell (1957) and popularized in terms of linear programming by Charnes, Cooper and Rhodes (1978). This approach is based on the assumptions that $\mathcal{S}$ is convex as well as free disposable. The economic implication of assuming $\mathcal{S}$ is convex is that average costs increase monotonically with output, which is termed *decreasing returns to scale* in the economic literature. The DEA estimator of $\mathcal{S}$ is defined as the convex-hull of $\hat{\mathcal{S}}_{\text{FDH}}$:

$$\begin{aligned}\hat{\mathcal{S}}_{\text{DEA}} \;=\; \Big\{&(x, y) \in \mathbb{R}^d \times \mathbb{R} \,:\, y \leq \sum_i \xi_i Y_i\,;\ x \geq \sum_i \xi_i X_i \ \text{ for some } \ (\xi_1, \xi_2, \ldots) \\ &\text{such that } \sum_i \xi_i = 1\,;\ \xi_i \geq 0,\ i = 1, 2, \ldots\Big\}.\end{aligned}$$

The DEA estimator of $\mathcal{S}$ is the set under the "lowest" concave and monotone function that lies above all points (X_i, Y_i)'s. The DEA estimator of the boundary g is then given by

$$\hat{g}_{\text{DEA}} = \max\{y \,:\, (x, y) \in \hat{\mathcal{S}}_{\text{DEA}}\}.$$

This estimator also has the same difficulty as the FDH estimator in its definition at a point x when there is no data smaller than x, but such probablity is negligible too in the limit under the statistical model considered in Section 3. The DEA estimator $\hat{\mathcal{S}}_{\text{DEA}}(\hat{g}_{\text{DEA}})$ is the nonparametric maximum likelihood estimator of $\mathcal{S}(g)$ on the class of all sets $\mathcal{S} \subset \mathcal{K}$ with monotone and concave g, provided (X_i, Y_i) are iid with the uniform distribution on $\mathcal{S}$.

When the free disposability assumption is dropped, the DEA approach may yield an inconsistent estimator. In this case, it is natural to use the convex-hull of the data set $\mathcal{P}$ as an estimator of the set $\mathcal{S}$, and to define the corresponding estimator $\hat{g}_{\text{conv}}$ of g in the same manner as that $\hat{g}_{\text{FDH}}$ and $\hat{g}_{\text{DEA}}$ are defined from $\hat{\mathcal{S}}_{\text{FDH}}$ and $\hat{\mathcal{S}}_{\text{DEA}}$, respectively. It is identical to the DEA estimator in the region where it is increasing. Also, its statistical properties are very similar to those of the DEA as is illustrated in Section 3.

2.3. The local polynomial estimators

Local polynomial methods are known to offer unsurpassed degree of flexibility and adaptivity in the context of regression and related problems. See for example Fan and Gijbels (1996). There is a variety of ways of applying them to density estimation, for example by using local likelihood techniques (Copas, 1995; Loader, 1996; Hjort and Jones, 1996; Park, Kim and Jones, 2000) or by converting the density estimation problem to one of regression (Park, Kim, Huh and Jeon, 1998). They have potentially a great deal to offer in problems of boundary estimation too. The first attempt was made by Hall, Park and Stern (1998), where the local polynomial boundary estimators were introduced for the case where $d = 1$. Multivariate extension of the local polynomial approaches shall be treated here. While the FDH and DEA approaches rely on the structural assumptions on $\mathcal{S}$, monotonicity and/or convexity, the local polynomial estimators are free of these restrictions.

2.3.1. The case where $d = 1$

First, we introduce the idea of local polynomial fitting for the case where $d = 1$. Multivariate extension shall be discussed later. Assume temporarily that the boundary is a polynomial of degree p, i.e. $g(x;\theta) = \theta_0 + \theta_1 x + \cdots + \theta_p x^p$, and that data are in the form of n independent random vectors (X_i, Y_i) having the uniform distribution on the region $\mathcal{S}(\theta) = \{(x,y) : 0 \le x \le 1,\ 0 \le y \le g(x;\theta)\}$. Here it is supposed that $\mathcal{D} = [0,1]$ and $g(x;\theta) > 0$ for all $x \in \mathcal{D}$ solely for simplicity of notation, but these assumptions are unimportant. One would then estimate $\theta = (\theta_0, \ldots, \theta_p)^T$ by maximizing the likelihood

$$\prod_{i=1}^{n} G(\theta)^{-1} I_{\mathcal{S}(\theta)}(X_i, Y_i)$$

where $G(\theta)$ is the area of $\mathcal{S}(\theta)$ given by $\int_0^1 g(x;\theta)\,dx$. The maximum likelihood estimator equals the value of θ that minimizes $G(\theta)$ subject to $0 \le Y_i \le g(X_i;\theta)$ for $1 \le i \le n$.

We may construct a nonparametric estimator of g by modifying the prescription in the previous paragraph. Suppose now that locally the boundary function g can be approximated by

$$g(u) \simeq \sum_{j=0}^{p} \theta_j (u-x)^j \tag{2.1}$$

with $\theta_j = g^{(j)}(x)/j!$ for u in a neighborhood of x. Let $\mathcal{I}_{x,h} = (x-h/2, x+h/2)$ for some $h > 0$. Then the approximation (2.1) is valid for $u \in \mathcal{I}_{x,h}$ with remainder $o(h^p)$ when g has p derivatives. Validity of the approximation (2.1) and the parametric considerations in the previous paragraph suggest the following local polynomial estimator:

$$\hat{g}_{\mathrm{LP}}(x) = \hat{\theta}_0$$

where $\hat{\theta} = (\hat{\theta}_0, \ldots, \hat{\theta}_p)^T$ minimizes

$$G(\theta; x, h) = \int_{\mathcal{I}_{x,h}} \{\theta_0 + \theta_1(u-x) + \cdots + \theta_p(u-x)^p\}\,du$$

subject to $Y_i \le \sum_{j=0}^{p} \theta_j (X_i - x)^j$ for all i: $X_i \in \mathcal{I}_{x,h}$.

The definition of the local likelihood estimator may be simplified by virtue of the fact that the first term, θ_0, in the approximating polynomial given at (2.1) does not involve u. To illustrate this, let κ_j equal 1 if j is even and 0 otherwise. We can write $G(\theta; x, h)/h = \theta_0 + \sum_{j=1}^{p} \kappa_j \theta_j (j+1)^{-1} 2^{-j} h^j$. Noting that $G(\theta; x, h)$ must be minimized subject to $\max^{(x)}\{Y_i - \theta_0 - \sum_{j=1}^{p} \theta_j (X_i - x)^j\} \leq 0$, we see that if $\hat{\theta}_1, \ldots, \hat{\theta}_p$ are known then

$$\hat{\theta}_0 = \max{}^{(x)} \left\{ Y_i - \sum_{j=1}^{p} \hat{\theta}_j (X_i - x)^j \right\} \tag{2.2}$$

where $\max^{(x)}$ denotes the maximum over i such that $X_i \in \mathcal{I}_{x,h}$. Define

$$H(\theta_1, \ldots, \theta_p) = \max{}^{(x)} \left[Y_i - \sum_{j=1}^{p} \theta_j \left\{ (X_i - x)^j - \kappa_j (j+1)^{-1} 2^{-j} h^j \right\} \right]. \tag{2.3}$$

It follows then that $G(\hat{\theta}_0, \hat{\theta}_1 \ldots, \hat{\theta}_p; x, h)/h = H(\hat{\theta}_1, \ldots, \hat{\theta}_p)$. Thus, $(\hat{\theta}_1, \ldots, \hat{\theta}_p)$ may be defined as the vector that minimizes H, and then $\hat{\theta}_0$ should be defined by (2.2).

2.3.2. Multivariate extension

Extension of the local polynomial boundary estimation to the case of multi-dimensional x is treated here. This requires careful notation to keep the expressions simple. We use the summation convention of Einstein: $a_i z^i = \sum_{i=1}^{d} a_i z_i$; $b_{ij} z^i z^j = \sum_{i=1}^{d} \sum_{j=1}^{d} b_{ij} z_i z_j$; $\cdots$. With this notation, whenever an index is repeated as a superscript and as a subscript, a summation over the range of that index is generated. Also, for an index set I we write $c_I z^I$ for the summations over all the indices in I. For example, if $I = \{i, j, k\}$ then $c_I z^I = c_{ijk} z^i z^j z^k = \sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{k=1}^{d} c_{ijk} z_i z_j z_k$.

Let $\theta_0 = g(x)$, $\theta_i = \partial g(x)/\partial x_i$, $\theta_{ij} = (1/2!) \partial^2 g(x)/\partial x_i \partial x_j$, and so on. Suppose that g is p-times continuously partially differentiable in a neighborhood of x. Then

$$g(u) = \theta_0 + \sum_{I:|I|=1}^{p} \theta_I (u - x)^I + o(|u - x|^p) \tag{2.4}$$

as $u \to x$, where $|I|$ denotes the size of the index set I. Define $\mathcal{R}_{x,h} = \Pi_{i=1}^{d} (x_i - h/2, x_i + h/2)$ for $h > 0$. The multivariate version of the local polynomial estimator is defined by $\hat{g}_{\mathrm{LP}}(x) = \hat{\theta}_0$ where $\hat{\theta}_0$, $\hat{\theta}_i$, $\hat{\theta}_{ij} (i \leq j)$, ... minimizes

$$G(\theta; x, h) = \int_{\mathcal{R}_{x,h}} \{\theta_0 + \sum_{I:|I|=1}^{p} \theta_I (u - x)^I\} \, du$$

subject to $Y_i \le \theta_0 + \sum_{I:|I|=1}^{p} \theta_I (X_i - x)^I$ for all i: $X_i \in \mathcal{R}_{x,h}$. It should be noted that the minimization is carried out with respect to θ_0, θ_i, $\theta_{ij}(i \le j)$, $\theta_{ijk}(i \le j \le k)$, ..., after taking $\theta_{ji} = \theta_{ij}$; $\theta_{ikj} = \theta_{jik} = \theta_{jki} = \theta_{kij} = \theta_{kji} = \theta_{ijk}$; etc.

To get analogues of (2.2) and (2.3), let $\mathcal{R} = (-1/2, 1/2)^d$, and define $m_i = \int_{\mathcal{R}} t_i \prod_{l=1}^{d} dt_l$, $m_{ij} = \int_{\mathcal{R}} t_i t_j \prod_{l=1}^{d} dt_l$, and so on. Note that there are many m's which are zero. For example, $m_i = 0$ for all i and $m_{ij} = 0$ for all $i \ne j$. However, for notational convenience we continue to write m_i, m_{ij}, ... instead of 0 in these cases. We may write then $G(\theta; x, h)/h^d = \theta_0 + \sum_{I:|I|=1}^{p} \theta_I m^I h^{|I|}$. Write θ_K for the vector formed by θ_i, $\theta_{ij}(i \le j)$, $\theta_{ijk}(i \le j \le k)$, ... It follows that $\hat{\theta}_K$ minimizes

$$H(\theta_K) = \max{}^{(x)} \left[Y_i - \sum_{I:|I|=1}^{p} \theta_I \left\{ (X_i - x)^I - m^I h^{|I|} \right\} \right] \tag{2.5}$$

where $\max^{(x)}$ denotes the maximum over i such that $X_i \in \mathcal{R}_{x,h}$, and that $\hat{\theta}_0$ is given by

$$\hat{\theta}_0 = \max{}^{(x)} \left\{ Y_i - \sum_{I:|I|=1}^{p} \hat{\theta}_I (X_i - x)^I \right\}. \tag{2.6}$$

3. Asymptotic Theory

Boundary estimation usually suffers from inherent bias that arises through having access only to data that lie on one side of the boundary. Thus, it is very important to be able to quantify this bias through explicit estimation and adjustment. An adequate solution to this problem requires at least some information about the distribution of the boundary estimator. In this section, the asymptotic distributions and the rates of convergence for the estimators introduced in the previous section are provided.

Some of the work on boundary estimation assumes Poisson-distributed points, and some assumes a given number, n, of independently distributed points. There is of course a duality between the two approaches, in which the intensity function of the former is replaced by n multiplied by the common probability density for the latter. First-order asymptotic results are generally the same in both contexts. We shall proceed here in the Poisson setting.

Let the data set $\mathcal{P} = \{(X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R} : i = 1, 2, \ldots\}$ be generated by a Poisson process with intensity $n\lambda$ in $\mathbb{R}^{d+1}$, where $n > 0$ is a scalar and $\lambda \ge 0$ is a fixed function on $\mathbb{R}^{d+1}$. Assume that $\lambda(x, y) = 0$ for $y > g(x)$, and that we observe those Poisson points (X_i, Y_i), for $i \ge 1$, that lie in a layer defined by

$\mathcal{S} = \{(x, y) : x \in \mathcal{D},\ g(x) - \varepsilon(x) < y < g(x)\}$, where ε is any positive function (possibly infinite) bounded away from zero on $\mathcal{D}$. Assume for some $\alpha > 0$ that λ satisfies, uniformly in $x \in \mathcal{D}$,

$$\lambda(x, y) = \alpha\{g(x) - y\}^{\alpha-1}\mu(x) + o[\{g(x) - y\}^{\alpha-1}] \tag{3.1}$$

as $y \uparrow g(x)$, where the function $\mu \geq 0$ is continuous. Call these assumptions $A(\alpha)$.

The assumptions $A(\alpha)$ have been considered in Härdle, Park and Tsybakov (1995), and Hall, Park and Stern (1998). Note that $\alpha - 1 > -1$ denotes the exponent of the rate at which λ decreases zero along the boundary. When $\alpha \leq 1$, we are in the case of a *sharp (fault-type) boundary*. When $\alpha > 1$, the intensity decreases to zero smoothly. In the case of the FDH estimator, asymptotic properties were derived by Park, Simar and Weiner (2000) for $d \geq 1$, but only the case $\alpha = 1$ is treated there. In fact, Park *et al.* (2000) established fully general asymptotic results in terms of measuring production efficiency, which includes the case where y is multivariate too. Asymptotic properties of the DEA estimator which are available now are due to Gijbels, Mammen, Park and Simar (1999), but they are very limited and only for the case where $d = 1$ and $\alpha = 1$. Theoretical properties of the local polynomial estimators were obtained by Hall, Park and Stern (1998) for the cases $d = 1$ and $\alpha > 0$. These existing asymptotic results for the three estimators shall be discussed in the next three subsections. Also, some new results on the rates of convergence for the FDH and the DEA, and on the asymptotic distributions of the local polynomial estimators for general $\alpha \geq 0$ and $d \geq 1$ shall be provided there.

3.1. The FDH estimator

In the case $\alpha = 1$, the assumption given at (3.1) reduces to that $\lambda(x, y) = \mu(x) + o(1)$ uniformly in $x \in \mathcal{D}$ as $y \uparrow g(x)$. Thus, $\lambda(x, g(x)) = \mu(x)$. Assume that g is continously partially differentiable, and that g is strictly increasing in each component, i.e. $\partial g(x)/\partial x_i > 0$ for each $i = 1, \ldots, d$. Call these conditions C_{FDH}. Write $g_1(x) = \Pi_{i=1}^d \partial g(x)/\partial x_i$. The following theorem demonstrates the asymptotic distribution of the FDH estimator, which is essentially due to Park, Simar and Weiner (2000).

Theorem 3.1. *Assume $A(1)$ and C_{FDH}. If $\mu(x) > 0$, then $n^{1/(d+1)}\{\hat{g}_{\text{FDH}}(x) - g(x)\}$ converges in law to a random variable Z_{FDH} whose distribution is given by*

$$G_{\text{FDH}}(z) = \exp[-\mu(x)(-z)^{d+1}/\{g_1(x)(d+1)!\}]$$

2.3.1. The case where $d = 1$

First, we introduce the idea of local polynomial fitting for the case where $d = 1$. Multivariate extension shall be discussed later. Assume temporarily that the boundary is a polynomial of degree p, i.e. $g(x;\theta) = \theta_0 + \theta_1 x + \cdots + \theta_p x^p$, and that data are in the form of n independent random vectors (X_i, Y_i) having the uniform distribution on the region $\mathcal{S}(\theta) = \{(x,y) : 0 \le x \le 1,\ 0 \le y \le g(x;\theta)\}$. Here it is supposed that $\mathcal{D} = [0,1]$ and $g(x;\theta) > 0$ for all $x \in \mathcal{D}$ solely for simplicity of notation, but these assumptions are unimportant. One would then estimate $\theta = (\theta_0, \ldots, \theta_p)^T$ by maximizing the likelihood

$$\prod_{i=1}^{n} G(\theta)^{-1} I_{\mathcal{S}(\theta)}(X_i, Y_i)$$

where $G(\theta)$ is the area of $\mathcal{S}(\theta)$ given by $\int_0^1 g(x;\theta)\,dx$. The maximum likelihood estimator equals the value of θ that minimizes $G(\theta)$ subject to $0 \le Y_i \le g(X_i;\theta)$ for $1 \le i \le n$.

We may construct a nonparametric estimator of g by modifying the prescription in the previous paragraph. Suppose now that locally the boundary function g can be approximated by

$$g(u) \simeq \sum_{j=0}^{p} \theta_j (u-x)^j \tag{2.1}$$

with $\theta_j = g^{(j)}(x)/j!$ for u in a neighborhood of x. Let $\mathcal{I}_{x,h} = (x - h/2, x + h/2)$ for some $h > 0$. Then the approximation (2.1) is valid for $u \in \mathcal{I}_{x,h}$ with remainder $o(h^p)$ when g has p derivatives. Validity of the approximation (2.1) and the parametric considerations in the previous paragraph suggest the following local polynomial estimator:

$$\hat{g}_{\mathrm{LP}}(x) = \hat{\theta}_0$$

where $\hat{\theta} = (\hat{\theta}_0, \ldots, \hat{\theta}_p)^T$ minimizes

$$G(\theta; x, h) = \int_{\mathcal{I}_{x,h}} \{\theta_0 + \theta_1(u-x) + \cdots + \theta_p(u-x)^p\}\,du$$

subject to $Y_i \le \sum_{j=0}^{p} \theta_j (X_i - x)^j$ for all i: $X_i \in \mathcal{I}_{x,h}$.

The definition of the local likelihood estimator may be simplified by virtue of the fact that the first term, θ_0, in the approximating polynomial given at (2.1) does not involve u. To illustrate this, let κ_j equal 1 if j is even and 0 otherwise. We can write $G(\theta; x, h)/h = \theta_0 + \sum_{j=1}^{p} \kappa_j \theta_j (j+1)^{-1} 2^{-j} h^j$. Noting that $G(\theta; x, h)$ must be minimized subject to $\max^{(x)}\{Y_i - \theta_0 - \sum_{j=1}^{p} \theta_j (X_i - x)^j\} \le 0$, we see that if $\hat\theta_1, \ldots, \hat\theta_p$ are known then

$$\hat\theta_0 = \max{}^{(x)} \left\{ Y_i - \sum_{j=1}^{p} \hat\theta_j (X_i - x)^j \right\} \tag{2.2}$$

where $\max^{(x)}$ denotes the maximum over i such that $X_i \in \mathcal{I}_{x,h}$. Define

$$H(\theta_1, \ldots, \theta_p) = \max{}^{(x)} \left[Y_i - \sum_{j=1}^{p} \theta_j \left\{ (X_i - x)^j - \kappa_j (j+1)^{-1} 2^{-j} h^j \right\} \right]. \tag{2.3}$$

It follows then that $G(\hat\theta_0, \hat\theta_1 \ldots, \hat\theta_p; x, h)/h = H(\hat\theta_1, \ldots, \hat\theta_p)$. Thus, $(\hat\theta_1, \ldots, \hat\theta_p)$ may be defined as the vector that minimizes H, and then $\hat\theta_0$ should be defined by (2.2).

2.3.2. Multivariate extension

Extension of the local polynomial boundary estimation to the case of multi-dimensional x is treated here. This requires careful notation to keep the expressions simple. We use the summation convention of Einstein: $a_i z^i = \sum_{i=1}^{d} a_i z_i$; $b_{ij} z^i z^j = \sum_{i=1}^{d} \sum_{j=1}^{d} b_{ij} z_i z_j$; $\cdots$. With this notation, whenever an index is repeated as a superscript and as a subscript, a summation over the range of that index is generated. Also, for an index set I we write $c_I z^I$ for the summations over all the indices in I. For example, if $I = \{i, j, k\}$ then $c_I z^I = c_{ijk} z^i z^j z^k = \sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{k=1}^{d} c_{ijk} z_i z_j z_k$.

Let $\theta_0 = g(x)$, $\theta_i = \partial g(x)/\partial x_i$, $\theta_{ij} = (1/2!)\partial^2 g(x)/\partial x_i \partial x_j$, and so on. Suppose that g is p-times continuously partially differentiable in a neighborhood of x. Then

$$g(u) = \theta_0 + \sum_{I:|I|=1}^{p} \theta_I (u - x)^I + o(|u - x|^p) \tag{2.4}$$

as $u \to x$, where $|I|$ denotes the size of the index set I. Define $\mathcal{R}_{x,h} = \Pi_{i=1}^{d}(x_i - h/2, x_i + h/2)$ for $h > 0$. The multivariate version of the local polynomial estimator is defined by $\hat g_{\mathrm{LP}}(x) = \hat\theta_0$ where $\hat\theta_0$, $\hat\theta_i$, $\hat\theta_{ij} (i \le j)$, ... minimizes

$$G(\theta; x, h) = \int_{\mathcal{R}_{x,h}} \{\theta_0 + \sum_{I:|I|=1}^{p} \theta_I (u - x)^I\}\, du$$

subject to $Y_i \le \theta_0 + \sum_{I:|I|=1}^{p} \theta_I (X_i - x)^I$ for all i: $X_i \in \mathcal{R}_{x,h}$. It should be noted that the minimization is carried out with respect to $\theta_0, \theta_i, \theta_{ij}(i \le j), \theta_{ijk}(i \le j \le k), \ldots$, after taking $\theta_{ji} = \theta_{ij}$; $\theta_{ikj} = \theta_{jik} = \theta_{jki} = \theta_{kij} = \theta_{kji} = \theta_{ijk}$; etc.

To get analogues of (2.2) and (2.3), let $\mathcal{R} = (-1/2, 1/2)^d$, and define $m_i = \int_{\mathcal{R}} t_i \prod_{l=1}^{d} dt_l$, $m_{ij} = \int_{\mathcal{R}} t_i t_j \prod_{l=1}^{d} dt_l$, and so on. Note that there are many m's which are zero. For example, $m_i = 0$ for all i and $m_{ij} = 0$ for all $i \neq j$. However, for notational convenience we continue to write m_i, m_{ij}, ... instead of 0 in these cases. We may write then $G(\theta; x, h)/h^d = \theta_0 + \sum_{I:|I|=1}^{p} \theta_I m^I h^{|I|}$. Write θ_K for the vector formed by θ_i, $\theta_{ij}(i \le j)$, $\theta_{ijk}(i \le j \le k)$, ... It follows that $\hat{\theta}_K$ minimizes

$$H(\theta_K) = \max{}^{(x)} \left[Y_i - \sum_{I:|I|=1}^{p} \theta_I \left\{ (X_i - x)^I - m^I h^{|I|} \right\} \right] \tag{2.5}$$

where $\max^{(x)}$ denotes the maximum over i such that $X_i \in \mathcal{R}_{x,h}$, and that $\hat{\theta}_0$ is given by

$$\hat{\theta}_0 = \max{}^{(x)} \left\{ Y_i - \sum_{I:|I|=1}^{p} \hat{\theta}_I (X_i - x)^I \right\}. \tag{2.6}$$

3. Asymptotic Theory

Boundary estimation usually suffers from inherent bias that arises through having access only to data that lie on one side of the boundary. Thus, it is very important to be able to quantify this bias through explicit estimation and adjustment. An adequate solution to this problem requires at least some information about the distribution of the boundary estimator. In this section, the asymptotic distributions and the rates of convergence for the estimators introduced in the previous section are provided.

Some of the work on boundary estimation assumes Poisson-distributed points, and some assumes a given number, n, of independently distributed points. There is of course a duality between the two approaches, in which the intensity function of the former is replaced by n multiplied by the common probability density for the latter. First-order asymptotic results are generally the same in both contexts. We shall proceed here in the Poisson setting.

Let the data set $\mathcal{P} = \{(X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R} : i = 1, 2, \ldots\}$ be generated by a Poisson process with intensity $n\lambda$ in $\mathbb{R}^{d+1}$, where $n > 0$ is a scalar and $\lambda \ge 0$ is a fixed function on $\mathbb{R}^{d+1}$. Assume that $\lambda(x, y) = 0$ for $y > g(x)$, and that we observe those Poisson points (X_i, Y_i), for $i \ge 1$, that lie in a layer defined by

$\mathcal{S} = \{(x, y) : x \in \mathcal{D},\ g(x) - \varepsilon(x) < y < g(x)\}$, where ε is any positive function (possibly infinite) bounded away from zero on $\mathcal{D}$. Assume for some $\alpha > 0$ that λ satisfies, uniformly in $x \in \mathcal{D}$,

$$\lambda(x, y) = \alpha\{g(x) - y\}^{\alpha-1}\mu(x) + o[\{g(x) - y\}^{\alpha-1}] \tag{3.1}$$

as $y \uparrow g(x)$, where the function $\mu \geq 0$ is continuous. Call these assumptions $A(\alpha)$.

The assumptions $A(\alpha)$ have been considered in Härdle, Park and Tsybakov (1995), and Hall, Park and Stern (1998). Note that $\alpha - 1 > -1$ denotes the exponent of the rate at which λ decreases zero along the boundary. When $\alpha \leq 1$, we are in the case of a *sharp (fault-type) boundary*. When $\alpha > 1$, the intensity decreases to zero smoothly. In the case of the FDH estimator, asymptotic properties were derived by Park, Simar and Weiner (2000) for $d \geq 1$, but only the case $\alpha = 1$ is treated there. In fact, Park *et al.* (2000) established fully general asymptotic results in terms of measuring production efficiency, which includes the case where y is multivariate too. Asymptotic properties of the DEA estimator which are available now are due to Gijbels, Mammen, Park and Simar (1999), but they are very limited and only for the case where $d = 1$ and $\alpha = 1$. Theoretical properties of the local polynomial estimators were obtained by Hall, Park and Stern (1998) for the cases $d = 1$ and $\alpha > 0$. These existing asymptotic results for the three estimators shall be discussed in the next three subsections. Also, some new results on the rates of convergence for the FDH and the DEA, and on the asymptotic distributions of the local polynomial estimators for general $\alpha \geq 0$ and $d \geq 1$ shall be provided there.

3.1. The FDH estimator

In the case $\alpha = 1$, the assumption given at (3.1) reduces to that $\lambda(x, y) = \mu(x) + o(1)$ uniformly in $x \in \mathcal{D}$ as $y \uparrow g(x)$. Thus, $\lambda(x, g(x)) = \mu(x)$. Assume that g is continously partially differentiable, and that g is strictly increasing in each component, i.e. $\partial g(x)/\partial x_i > 0$ for each $i = 1, \ldots, d$. Call these conditions C_{FDH}. Write $g_1(x) = \Pi_{i=1}^{d} \partial g(x)/\partial x_i$. The following theorem demonstrates the asymptotic distribution of the FDH estimator, which is essentially due to Park, Simar and Weiner (2000).

Theorem 3.1. *Assume $A(1)$ and C_{FDH}. If $\mu(x) > 0$, then $n^{1/(d+1)}\{\hat{g}_{\text{FDH}}(x) - g(x)\}$ converges in law to a random variable Z_{FDH} whose distribution is given by*

$$G_{\text{FDH}}(z) = \exp[-\mu(x)(-z)^{d+1}/\{g_1(x)(d+1)!\}]$$

for $z < 0$ *and* $G_{\mathrm{FDH}}(z) = 1$ *for* $z \geq 0$.

It follows from the theorem that asymptotically $n^{1/(d+1)}\{g(x) - \hat{g}_{\mathrm{FDH}}(x)\}$ follows a Weibull distribution with $\{g_1(x)(d+1)!/\mu(x)\}^{1/(d+1)}$ as a scale, and $(d+1)$ as a shape parameter. Since the r-th moment of a standard Weibull distribution with a shape parameter c is given by $\Gamma((r+c)/c))$, the asymptotic r-th moment of $\hat{g}_{\mathrm{FDH}}(x)$ is

$$(-1)^r n^{-r/(d+1)} \left\{\frac{g_1(x)(d+1)!}{\mu(x)}\right\}^{r/(d+1)} \Gamma\left(\frac{d+r+1}{d+1}\right).$$

It may be proved that, under the assumptions $A(\alpha)$, instead of $A(1)$, the FDH estimator has the convergence rate $n^{-1/(d+\alpha)}$ for boundaries that satisfy a Lipschitz condition of order 1. In fact, this convergence rate is minimax-optimal in a pointwise sense under the Lipschitz condition, which may be proved as in Härdle, Park and Tsybakov (1995). We point out that existence and continuity of the partial derivatives which together are stronger than the Lipschitz condition are imposed in Theorem 3.1 only so that the limit distribution might be identified.

Statistical properties of $\hat{\mathcal{S}}_{\mathrm{FDH}}$ as a set estimator of $\mathcal{S}$ when $\mathcal{S}$ is compact have been also studied. We refer to Korostelev, Simar and Tsybakov (1995a,b) for those results. We close this subsection by pointing out that if one of the partial derivatives $\partial g(x)/\partial x_i$ is zero then $\hat{g}_{\mathrm{FDH}}(x)$ converges to $g(x)$ faster than $n^{-1/(d+1)}$.

3.2. The DEA estimator

Asymptotic distribution of the DEA estimator is available only for the case where $d = 1$ and $\alpha = 1$. The following theorem is due to Gijbels, Mammen, Park and Simar (1999). To state the theorem, we assume that g is twice continuously differentiable, monotone and strictly concave, i.e. $g''(x) < 0$. Call these assumptions C_{DEA}. Write $g_2(x) = -g''(x)/2$. For $z < 0$ and $v > 0$, let $h(v, z; x) = (1/2)\mu(x)\{g_2(x)v^2 - z\}\exp[-(1/6)\mu(x)g_2(x)^{-2}v^{-3}\{g_2(x)v^2 - z\}^3]$.

Theorem 3.2. *Assume* $A(1)$ *and* C_{DEA}. *If* $\mu(x) > 0$, *then* $n^{2/3}\{\hat{g}_{\mathrm{DEA}}(x) - g(x)\}$ *converges in law to a random variable* Z_{DEA} *whose distribution is given by*

$$G_{\mathrm{DEA}}(z) = \int_0^\infty h(v, z; x)\, dv$$

for $z < 0$ *and* $G_{\mathrm{DEA}}(z) = 1$ *for* $z \geq 0$.

The asymptotic distribution function $G_{\rm DEA}(z)$ is continuous at $z=0$. To see this, note that for $-\epsilon < z < 0$ the integrand $h(v,z;x)$ is bounded by

$$(1/2)\mu(x)\{g_2(x)v^2+\epsilon\}\exp\{-(1/6)\mu(x)g_2(x)v^3\}$$

which is integrable over $v \in (0,\infty)$. Thus, $\lim_{z\uparrow 0} G_{\rm DEA}(z) = \int_0^\infty \lim_{z\uparrow 0} h(v,z;x)\,dv = 1$.

One may obtain a simplified expression for the limit distribution which is convenient for calculating the asymptotic moments. To find it, make transformation: $v \mapsto u$ by $v = \{-z/g_2(x)\}^{1/2}u$ for the integral at Theorem 3.1. By a simple algebraic manipulation one may see that $n^{2/3}\{\mu(x)^2/g_2(x)\}^{1/3}\{\hat{g}_{\rm DEA}(x) - g(x)\}$ converges in distribution to a random variable $W_{\rm DEA}$ whose distribution function is given by

$$P(W_{\rm DEA} \le z) = \int_0^\infty \varphi(u,z;x)\,du$$

for $z<0$, where $\varphi(u,z;x) = (1/2)(-z)^{3/2}(1+u^2)\exp\{-(1/6)(-z)^{3/2}(u+u^{-1})^3\}$. The r-th moment of $W_{\rm DEA}$ is easily obtained by an elementary calculation, from which the r-th asymptotic moment of $\hat{g}_{\rm DEA}(x)$ may be derived as

$$(-1)^r n^{-2r/3}\, 4\cdot 6^{(2r/3)-1}\left\{\frac{g_2(x)}{\mu(x)^2}\right\}^{r/3} k\,\Gamma\left(\frac{2k}{3}\right)\frac{k!(k+1)!}{(2k+1)!}.$$

The asymptotic distribution of the DEA estimator for multivariate x is unknown. However, its rate of convergence was found by Kneip, Park and Simar (1998) in very general statistical models which include the case of multivariate x. When $\alpha=1$ and x is d-dimensional, it was shown that $\hat{g}_{\rm DEA}(x)$ converges to $g(x)$ at the rate $n^{-2/(d+2)}$ under the condition that g satisfies a Lipschitz condition of order 1 on its first derivative. (In fact, it was also shown that the rate can not be improved further by imposing more smoothness assumptions on g.) This result may be extended to the case where $\alpha > 0$. It can be shown that the DEA estimator achieves the convergence rate $n^{-2/(d+2\alpha)}$ under $A(\alpha)$ and the Lipschitz condition on the boundaries. The rate is minimax-optimal in a pointwise sense, which may be proved by the techniques used in Härdle, Park and Tsybakov (1995).

A closely related estimator, $\hat{g}_{\rm conv}$ (the convex-hull estimator) mentioned in Subsection 2.2, may be shown to have the same limit law as $\hat{g}_{\rm DEA}$ under weaker conditions where monotonicity of g drops out. Related work on the convex-hull method includes that of Nagaev (1995) on properties of the convex-hull in the case where the Poisson point process has an unbounded convex domain, and that of Rényi and Sulanke (1963, 1964), Efron (1965), Groeneboom (1988), and Cabo and Groeneboom (1994) on the number of vertices (and other quantities) of the convex-hull of random points. A generalization of the convex-hull method in boundary estimation was treated by Hall, Park and Turlach (2001).

3.3. The local polynomial estimator

Asymptotic properties of the local polynomial estimators were investigated by Hall, Park and Stern (1998) for the case where $d = 1$ and $\alpha > 0$. Here, they are generalized to the case where $d \geq 1$. The p-th order local polynomial estimators defined as at (2.2) and (2.6) are considered. Assume g has $p+1$ continuous partial derivatives. Let h be asymptotic to $Cn^{-1/\{d+\alpha(p+1)\}}$ for some constant $C > 0$. Call these conditions C_{LP}. The size of the bandwidth optimizes the convergence rate of $\hat{g}_{\mathrm{LP}}$ to g.

To state the theorem, define $\omega = \mu(x)^{1/\alpha}C^{(d/\alpha)+p+1}$. Let $W_1, W_2, \ldots$ be independent random variables which are exponentially distributed with density e^{-w} for $w > 0$. Let γ denote Euler's constant, i.e. $\gamma = \lim_{k\to\infty}\{\sum_{j=1}^{k-1}(1/j) - \log k\}$. Define for $j \geq 1$

$$U_r = \exp\left[-\alpha^{-1}\left\{\sum_{i=r}^{\infty} i^{-1}(W_i - 1) + \gamma - \sum_{i=1}^{r-1} i^{-1}\right\}\right].$$

With this definition $(U_1, U_2, \ldots)$ obey the ordering $0 \leq U_1 \leq U_2 \leq \cdots$, and have the joint distribution of consecutive Type 2 extreme values in the sense of Gnedenko (1943). See Hall (1978).

Next, let $V_1, V_2, \ldots$ be independent random variables, independent of $W_1, W_2, \ldots$, having the uniform distribution on $[-1/2, 1/2]^d$. Let d_i, d_{ij}, d_{ijk}, $\ldots$ be sequences with multiple indices such that $d_{ji} = d_{ij}$; $d_{ikj} = d_{jik} = d_{jki} = d_{kij} = d_{kji} = d_{ijk}$; etc, where $1 \leq i \leq d$, $1 \leq j \leq d$, $1 \leq k \leq d$, $\ldots$. Let d_K be the vector formed by d_i, $d_{ij}(i \leq j)$, $d_{ijk}(i \leq j \leq k)$, $\ldots$. Define

$$S(d_K) = \inf_{1\leq r<\infty}\left\{U_r + \sum_{I:|I|=1}^{p} d_I\left(V_r^I - m^I\right) - \omega\theta_{I_{p+1}}V_r^{I_{p+1}}\right\}$$

where I_{p+1} denotes an index set of size $p+1$. Let D_K denote the value of d_K

that maximizes $S(d_K)$. Define

$$D_0 = -\inf_{1\le r<\infty}\left(U_r + \sum_{I:|I|=1}^{p} D_I V_r^I - \omega\theta_{I_{p+1}} V_r^{I_{p+1}}\right).$$

Theorem 3.3. *Assume $A(\alpha)$ and C_{LP}. If $\mu(x)>0$, then*

$$\{C^d\mu(x)\}^{1/\alpha} n^{(p+1)/\{d+\alpha(p+1)\}}\{\hat{g}_{\mathrm{LP}}(x)-g(x)\}\to D_0$$

in distribution.

The theorem may be proved as in Hall, Park and Stern (1998). The convergence rate $n^{-(p+1)/\{d+\alpha(p+1)\}}$ evinced by Theorem 3.3 is minimax-optimal in a point pointwise sense for boundaries that have p derivatives and satisfy a Lipschitz condition of order 1 on the p-th derivative. This can be also proved as in Härdle, Park and Tsybakov (1995).

The representation for the limit law of the local polynomial estimators presented in Theorem 3.3 enables computation of numerical approximations to the moments of the asymptotic distribution. The result may be applied directly to correct for bias, or more generally to compute empirical, bias-adjusted, confidence intervals for the unknown boundary.

4. Other Related Works and Future Research

Bias correction is a particularly important problem in boundary estimation. This problem was tackled by Park, Simar and Weiner (1998) for the FDH, Gijbels, Mammen, Park and Simar (1999) for the DEA, and Hall, Park and Stern (1998) for the local polynomial estimators. But, the methods of correcting for bias considered there are restrictive and specialized to those estimators. A rather general approach that has versions for any boundary estimator was suggested by Hall and Park (2000). It is a new form of the bootstrap, termed the *translation bootstrap*, involving averaging over repeated empirical translations. It should be noted that the usual bootstrap procedure does not produce even asymptotically consistent results since it is unable to accurately capture the relationship among extremes of a resample drawn by resampling in the usual way. The subsampling bootstrap approaches of Bickel, Götze and van Zwet (1997) and Politis, Romano and Wolf (1999) would work, but for effective performance they require empirical choice of subsample size.

Open issues for future research include (i) searching for asymptotic distribution of the DEA estimator when $d > 1$ and/or $\alpha > 0$ and that of the FDH when $\alpha > 0$; (ii) bandwidth selection for the local polynomial approaches; (iii) developing theory for estimating derivatives of a boundary; (iv) bias correction.

Extension of Theorem 3.1 to the case where $\alpha > 0$ does not seem difficult, but that of Theorem 3.2 to the multivariate case appears to be a very difficult problem because of the level of complexity it requires. Regarding the second issue, the well-known techniques for choosing bandwidth in conventional curve estimation problems include plug-in rules, cross-validation and the bootstrap. In the context of boundary estimation plug-in approaches are not promising since one can hardly expect explicit expression for the multiplicative constant in the formula for the asymptotically optimal bandwidth. The idea of cross-validation seems also difficult to implement since there exists no empirical substitute for the cross-product contribution to the integrated squared error. Bootstrap procedures with relevant resampling schemes or the translation bootstrap techniques of Hall and Park (2000) might be successful.

Estimating derivatives of a boundary is also an important problem. Recent work on derivative estimation includes that of Park (2001). As in regression and density settings, the problem arises when one wants to correct for bias and to construct empirical confidence intervals for the unknown data edge. It is an interesting problem in its own right due to some important interpretations of boundary derivatives in economic applications.

References

Bickel, P., Götze, F. and van Zwet, W.R. (1995). Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistica Sinica* 7, 1–31.

Cabo, A.J. and Groeneboom, P. (1994). Limit theorems for functionals of convex hulls. *Probability Theory and Related Fields* 100, 31–55.

Charnes, A., Cooper, W.W. and Rhodes, E. (1978). Measuring the inefficiency of decision making units. *European Journal of Operational Research* 2, 429–444.

Copas, J.B. (1995). Local likelihood based on kernel censoring. *Journal of the Royal Statistical Society, Series B* 57, 221–235.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica* 19, 273–292.

Deprins, D., Simar, L. and Tulkens, H. (1984). Measuring labor inefficiency in post offices. In M. Marchand, P. Pestieau and H. Tulkens (eds.) *The Performance of Public Enterprises: Concepts and Measurements.* North-Holland, Amsterdam.

Efron, B. (1965). The convex hull of a random set of points. *Biometrika* 52, 331–343.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

Farrell, M.J. (1957). The measurement of productivity efficiency. *Journal of the Royal Statistical Society, Series A* 120, 253–281.

Gijbels, I., Mammen, E., Park, B.U. and Simar, L. (1999). On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association* 94, 220–228.

Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *The Annals of Mathematics* 44, 423–453.

Greene, W.H. (1980). Maximum likelihood estimation of econometric frontier. *Journal of Econometrics* 13, 27–56.

Groeneboom, P. (1988). Limit theorems for convex hulls. *Probability Theory and Related Fields* 79, 327–368.

Härdle, W., Park, B.U. and Tsybakov, A.B. (1995). Estimation of non-sharp support boundaries. *Journal of Multivariate Analysis* 55, 205–218.

Hall, P. (1978). Representations and limit theorems for extreme value distributions. *Journal of Applied Probability* 15, 639–644.

Hall, P. and Park, B.U. (2000). New methods for bias correction at endpoints and boundaries. *The Annals of Statistics*, under revision.

Hall, P., Park, B.U. and Stern, S.E. (1998). On polynomial estimators of frontiers and boundaries. *Journal of Multivariate Analysis* 66, 71–98.

Hall, P., Park, B.U. and Turlach, B.A. (2001). Rolling-ball method for estimating the boundary of the support of a point-process intensity. *Annales de l'Institut Henri Poincare-Probabilites et Statistiques*, to appear.

Hausman, J.A. and Taylor, W.E. (1981). Panel data and unobservable individual effects. *Econometrica* 49, 1377–1398.

Hjort, N.L. and Jones, M.C. (1996). Locally parametric nonparametric density estimation. *The Annals of Statistics* 24, 1619–1647.

Kneip, A., Park, B.U. and Simar, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory* 14, 783–793.

Koopmans, T.C. (1951). An analysis of production as an efficient combination of activities. In T.C. Koopmans (ed.), *Activity Analysis of Production and Allocation*. Cowles Commission for Research in Economics, Monograph 13. John-Wiley, New York.

Korostelev, A.L., Simar, L. and Tsybakov, A.B. (1995a). Efficient estimation of monotone boundaries. *The Annals of Statistics* 23, 476–489.

Korostelev, A.L., Simar, L. and Tsybakov, A.B. (1995b). On estimation of monotone and convex boundaries. *Pub. Inst. Statist. Univ. Paris* 49, 3–18.

Korostelev, A.L. and Tsybakov, A.B. (1993). *Minimax Theory of Image Reconstruction*, Lecture Notes in Statistics (82). Springer, New York.

Loader, C.R. (1996). Local likelihood density estimation. *The Annals of Statistics* 24, 1602–1618.

Nagaev, A.V. (1995). Some properties of convex hulls generated by homogeneous Poisson point processes in an unbounded convex domain. *Ann. Inst. Statist. Math* 47, 21–29.

Park, B.U. (2001). On estimating the slope of increasing boundaries. *Statistics and Probability Letters* 52, 69–72.

Park, B.U., Kim, W.C. and Jones, M.C. (2000). On local likelihood density estimation. Submitted to *The Annals of Statistics*.

Park, B.U., Kim, W.C., Huh, J. and Jeon, J.W. (1998). Estimation of density via local polynomial regression. *Journal of the Korean Statistical Society* 27, 91–100.

Park, B.U., Sickles, R.C. and Simar, L. (1998). Stochastic frontiers: A semiparametric approach. *Journal of Econometrics* 84, 273–301.

Park, B.U., Sickles, R.C. and Simar, L. (2000). Semiparametric efficient estimation of AR(1) panel data models. Submitted to *Journal of Econometrics.*

Park, B.U. and Simar, L. (1994). Efficient semiparametric estimation in stochastic frontier models. *Journal of the American Statistical Association* 89, 929–936.

Park, B.U., Simar, L. and Weiner, Ch. (2000). The FDH estimator for productivity efficiency scores. *Econometric Theory* 16, 855–877.

Politis, D., Romano, J.P. and Wolf, M. (1999). *Subsampling.* Springer, New York.

Rényi, A. and Sulanke, R. (1963). On the convex hull of n randomly chosen points. *Z. Wahrscheinlichkeitstheorie verw. Geb.* 2, 75–84.

Rényi, A. and Sulanke, R. (1964). On the convex hull of n randomly chosen points II. *Z. Wahrscheinlichkeitstheorie verw. Geb.* 3, 138–147.

Shephard, R.W. (1970). *Theory of Cost and Production Function.* Princeton University Press, Princeton.