

Tripathi, Gautam; Kim, Woocheol

Working Paper

Nonparametric estimation of homogeneous function

SFB 373 Discussion Paper, No. 2000,85

Provided in Cooperation with:

Collaborative Research Center 373: Quantification and Simulation of Economic Processes,
Humboldt University Berlin

Suggested Citation: Tripathi, Gautam; Kim, Woocheol (2000) : Nonparametric estimation of homogeneous function, SFB 373 Discussion Paper, No. 2000,85, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, <https://nbn-resolving.de/urn:nbn:de:kobv:11-10048116>

This Version is available at:

<https://hdl.handle.net/10419/62177>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

NONPARAMETRIC ESTIMATION OF HOMOGENEOUS FUNCTIONS

GAUTAM TRIPATHI AND WOOCHEOL KIM

ABSTRACT. Consider the regression $y = f(\tilde{x}) + \varepsilon$ where $\mathbb{E}(\varepsilon|\tilde{x}) = 0$ and the exact functional form of f is unknown, although we do know that it is homogeneous of known degree r . Using a local linear approach we examine two ways of nonparametrically estimating f : (i) a “direct” or “numeraire” approach, and (ii) a “projection based” approach. We show that depending upon the nature of the conditional variance $\text{var}(\varepsilon|\tilde{x})$, one approach may be asymptotically better than the other. Results of a small simulation experiment are presented to support our findings.

1. INTRODUCTION

An important problem in microeconometrics is the estimation of shape restricted functions. In order to obtain good estimates without worrying about any potential misspecification problems, imposing valid shape restrictions on nonparametric estimators of these functional forms seems like a good idea. Starting from the pioneering paper of Hildreth (1954) much work has been done in this area. Some recent references include, among others, Gallant (1981), Yatchew (1988), Härdle (1989, Chapter 8), Ryu (1993), Matzkin (1994), Ruud (1997) and Yatchew and Bos (1997). For readers unfamiliar with techniques of nonparametric estimation relevant to econometrics, we recommend Bierens (1985), Härdle (1989), Härdle and Linton (1994), Yatchew (1998), and Pagan and Ullah (1999).

In this paper we restrict ourselves to estimating a conditional mean function f which happens to be homogeneous of known degree r . Recall that $f : S \subset \mathbb{R}^s \rightarrow \mathbb{R}$ is said to be homogeneous of degree $r \in \mathbb{R}$ on S iff $f(\lambda\tilde{x}) = \lambda^r f(\tilde{x})$ for all $(\lambda, \tilde{x}) \in \mathbb{R}_{++} \times S$ such that $\lambda\tilde{x} \in S$. Such functional

Date: This Version: February 21, 2000.

Key words: Homogeneous Functions, Locally Linear Estimators, Nonparametric Estimation.

Address: Gautam Tripathi: Department of Economics, University of Wisconsin, Madison, WI-53706, USA. *E-Mail:* gtripath@ssc.wisc.edu. Woocheol Kim: Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin, D-10178 Berlin, Germany. *E-mail:* woocheol@wiwi.hu-berlin.de.

The first author would like to thank the hospitality of Professor Wolfgang Härdle at the Institute of Statistics and Econometrics, Humboldt University, Berlin, where part of this research was carried out. Financial support to the first author from Sondersforschungsbereich 373, “Quantifikation und Simulation Ökonomischer Prozesse,” is also gratefully acknowledged.

forms are frequently encountered in microeconomic theory. For instance, the profit (resp. cost) function for a profit maximizing (resp. cost minimizing) competitive firm is homogeneous of degree one in prices. Similarly, the Marshallian demand functions for a utility maximizing agent are homogeneous of degree zero in prices and income. In production theory attention is often restricted to production functions which are homogeneous of degree one; i.e. which exhibit constant returns to scale. See, for instance, the classic paper by Arrow, Chenery, Minhas, and Solow (1961).

Although many functional forms familiar to economists may satisfy other shape restrictions besides homogeneity, for now we focus upon homogeneity alone. One reason for doing so is that, compared with some other shape properties such as concavity or monotonicity, homogeneity is a particularly tractable property to analyze. Loosely speaking, this is because the set of all homogeneous functions (embedded in some larger space such as the set of all twice continuously differentiable functions) is a linear space. This linearity allows for ease of analysis in many situations. On the other hand, the set of all concave, or monotone, functions is not a linear space but a convex subset of the ambient space. Typically, this makes dealing with concave, or monotone, functions more difficult. Therefore, focusing on homogeneity alone may often lead to a simplification of econometric analysis. Furthermore, as a practical matter, imposing concavity and monotonicity restrictions on function estimates seems to be a hard, although not an impossible, task. On the other hand, imposing homogeneity in nonparametric estimates is quite easy and may lead to substantial improvement of estimates in finite samples.

In the parametric case it is well known how to impose a homogeneity restriction. Basically, the idea is to restrict the parameter space. For example, in a log linear Cobb-Douglas regression model with two covariates, homogeneity is imposed by requiring that the coefficients on the two factors sum to one. Even in the flexible functional form literature homogeneity is imposed by restricting the parameter space. For instance, Gallant (1981) imposes constant returns to scale by making some parameters in a Fourier flexible form expansion sum to unity. Slightly differently, Ryu (1993) shows how to impose linear homogeneity by a polar coordinate transformation.

However, in the fully nonparametric case perhaps the simplest way of imposing homogeneity is to use a “direct” or “numeraire” approach. In this approach we pick one variable as the numeraire, and use it to normalize all variables. Estimation is then carried out using these normalized variables. For instance, Ruud (1997, Page 171) follows this approach in imposing homogeneity on his shape restricted estimator. From our conversations with many colleagues we get the impression that most economists immediately think of the numeraire approach when asked to nonparametrically estimate a homogeneous conditional expectation. Since we are so used to working with ratios of variables such as relative prices etc., which are homogeneous of degree zero by construction, the choice of the “direct” approach is perhaps not very surprising.

But there is another way of nonparametrically estimating homogeneous conditional means. We call this the “projection based” approach for reasons that will be clarified later on. In this paper we show how to implement the “projection based” and “direct” approaches using local linear estimators and compare the asymptotic properties of the estimates obtained. Their analytical simplicity and ease of use should make the proposed estimators a useful addition to the toolkit of the applied econometrician.

The paper is organized as follows: Section 2 lists the maintained assumptions and Section 3 describes the procedure for estimating homogeneous functions using the “direct” and the “projection based” approach. The estimators are implemented using the local linear approach. In Section 4 we compare the asymptotic performance of the “direct” and “projection based” estimators and show that the nature of the conditional variance of the error term determines which estimator is better. Section 5 describes the results of a small simulation experiment and Section 6 concludes. All proofs are confined to the appendices.

Notation 1.1. The following notation is used throughout this paper. We treat all vectors as column vectors and (most of the time) denote them explicitly by using a tilde. In particular, $\tilde{x} = (x_1, \dots, x_s)$, $\tilde{w} = (\frac{x_1}{x_s}, \dots, \frac{x_{s-1}}{x_s})$, $\tilde{x}_j = (x_{1,j}, \dots, x_{s,j})$, and $\tilde{w}_j = (\frac{x_{1,j}}{x_{s,j}}, \dots, \frac{x_{s-1,j}}{x_{s,j}})$. $S_{\tilde{x}}^b$ is a convex compact subset of \mathbb{R}^s such that x_s (the last component of $\tilde{x} \in S_{\tilde{x}}^b$) is positive and bounded away from zero. Let $H_0 : S_{\tilde{x}}^b \rightarrow \mathbb{R}^{s-1}$ be the homogeneous of degree zero transformation $H_0(\tilde{x}) = \tilde{w}$, and let $S_{\tilde{w}} = H_0(S_{\tilde{x}}^b)$. $C^k(int(S_{\tilde{x}}^b))$ is the set of all real valued functions on $int(S_{\tilde{x}}^b)$ which have continuous partial derivatives up to order k . We say that $f \in C^k(S_{\tilde{x}}^b)$ if $f \in C^k(int(S_{\tilde{x}}^b))$ and f , including all its partial derivatives up to order k , can be extended continuously to $S_{\tilde{x}}^b$. $L^2(S_{\tilde{x}}^b)$ is the set of all square integrable functions on $S_{\tilde{x}}^b$ which are integrable w.r.t the probability distribution on $S_{\tilde{x}}^b$. \mathcal{F}_r (resp. \mathcal{G}_r) is the set of all functions in $C^2(S_{\tilde{x}}^b)$ (resp. $L^2(S_{\tilde{x}}^b)$) which are also homogeneous of degree r . The symbol “ $\stackrel{\approx}{=}$ ” indicates “approximate equality”; i.e. equality modulo an additive, but asymptotically negligible, term. \square

2. THE SETUP

Consider the nonparametric regression $y_j = f(\tilde{x}_j) + \varepsilon_j$.

Assumption 2.1. *The following assumptions are maintained:*

- (i) *The data $\{y_j, \tilde{x}_j\}_{j=1}^n$ are iid random variables in $\mathbb{R} \times S_{\tilde{x}}^b$ and $\mathbb{E}(\varepsilon|\tilde{x}) = 0$.*
- (ii) *The functional form of $f \in \mathcal{F}_r$ is unknown but we do know r .*
- (iii) *Conditional pdf of $(y, x_s|\tilde{w})$ is twice continuously differentiable in \tilde{w} .*
- (iv) *$h \in C^2(S_{\tilde{w}})$ is the pdf of $\tilde{w} = H_0(\tilde{x})$ and $\inf_{\tilde{w} \in S_{\tilde{w}}} h(\tilde{w}) > 0$.*
- (v) *For some $\delta > 0$ the map $\tilde{w} \mapsto \mathbb{E}(|\varepsilon x_s^r|^{2+\delta} | \tilde{w})$ is bounded and continuous on $S_{\tilde{w}}$.* \square

The restrictions on $S_{\tilde{x}}^b$, namely that $S_{\tilde{x}}^b$ is compact and that the last coordinate of $\tilde{x} \in S_{\tilde{x}}^b$ is positive and bounded away from zero, ensure that

conditional expectations occurring in the statements of Lemmas A.1 and A.2 exist for all $r \in \mathbb{R}$. This allows us to handle any degree of homogeneity. Since $S_{\tilde{x}}^b$ is compact, (ii) implies that f is also an element of \mathcal{G}_r . (iii) implies that $\mathbb{E}(yx_s^r|\tilde{w})$, $\mathbb{E}(y^2x_s^{2r}|\tilde{w})$ and $\mathbb{E}(yx_s^{3r}|\tilde{w})$ are twice continuously differentiable in \tilde{w} . This fact is used in the proof of Lemma B.1. We use (iv) to ensure that the remainder terms in the Taylor expansions employed in Appendix A are well behaved. This assumption is regularly invoked in the kernel estimation literature to minimize complexity of mathematical details when dealing with ratios of random variables¹. See Newey (1994, Page 242) for a brief discussion on this type of assumption. As Newey points out, it is possible to relax this assumption but at the expense of a considerable increase in mathematical detail. (iv) can be made more palatable if we interpret it to mean that we should carry out estimation and inference in a region where the density is indeed bounded away from zero. The last assumption provides sufficient moments so that we can prove the asymptotic normality of estimators of f . See Lemma A.1.

3. ESTIMATION HEURISTICS

Since f is homogeneous of degree r we can write

$$(3.1) \quad y = x_s^r f(\tilde{w}, 1) + \varepsilon \iff \frac{y}{x_s^r} = f(\tilde{w}, 1) + \frac{\varepsilon}{x_s^r}.$$

The problem we investigate in this paper can be stated quite simply: Should we estimate $f(\tilde{x})$ using the first representation or the second? We refer to the estimate of $f(\tilde{x})$ based on the first representation as a “projection based” estimate, while the estimate based on the second formulation is called a “direct” or “numeraire” estimate. It may not be very obvious at this point how we can estimate $f(\tilde{x})$ using the first representation but, as we shall soon show, it is quite easy to do so. Although algebraically equivalent, the two formulations will in general lead to estimators with different statistical properties since division by x_s^r alters the stochastic properties of the error term ε . In fact, and this should not surprise the reader, the statistical performance of the estimators depends upon the conditional variance $\text{var}(\varepsilon|\tilde{x})$. In particular, we will show that if $\text{var}(\varepsilon|\tilde{x})$ is homogeneous of degree zero (which implies homoscedasticity as a special case) then the “projection based” estimator is asymptotically better than the “direct” estimator, while if $\text{var}(\varepsilon|\tilde{x})$ is homogeneous degree $r \neq 0$ then the latter dominates the former.

Remark 3.1. Notice that if $r = 0$, i.e. if we are estimating a homogeneous of degree zero function, for example a demand function, the two approaches will yield identical results. Furthermore, if $s = 1$ the problem is uninteresting since homogeneous functions are known up to scale in the one dimensional

¹However, it seems quite restrictive. For instance, if $\tilde{x} = (x_1, x_2)$ where $x_1, x_2 \stackrel{d}{\sim}$ UIID[1, 2] then it is straightforward to verify that the pdf of $w = x_1/x_2$ vanishes at the boundary and is not differentiable at $w = 1$; i.e. Assumption (iv) rules out this case.

case (because when $s = 1$, homogeneity of f implies that $f(x) = x^r f(1)$). To avoid these trivial cases from now on we assume that $r \neq 0$ and $s > 1$. \square

Our estimation strategy is to approximate sample analogs of optimization problems that identify $f(\tilde{x})$ using the two representations in (3.1). At the population level we can use the first representation to write the “projection based” specification of $f(\tilde{x})$ as $x_s^r \beta_p(\tilde{x})$, where β_p is identified as

$$(3.2) \quad \beta_p = \underset{\{\beta: \beta \in L^2(S_{\tilde{w}})\}}{\operatorname{argmin}} \mathbb{E}\{y - x_s^r \beta(\tilde{w})\}^2.$$

Since $x_s^r \beta(\tilde{w})$ is a homogeneous function of degree r for all $\beta \in L^2(S_{\tilde{w}})$, we can characterize $x_s^r \beta_p(\tilde{w})$ as the orthogonal projection of y onto \mathcal{G}_r under the “usual” L^2 inner product $\langle u, v \rangle_{L^2} = \mathbb{E}\{uv\}$. In particular, we can use the result in Tripathi (1999) to show that this projection can be explicitly calculated as $\frac{x_s^r \mathbb{E}(yx_s^r | \tilde{w})}{\mathbb{E}(x_s^{2r} | \tilde{w})}$. This explains our usage of the term “projection based” in describing an estimate obtained by using them first representation in (3.1). Similarly, a population level specification of $f(\tilde{x})$ using the second representation can be written as $x_s^r \beta_d(\tilde{w})$, where we identify β_d as

$$(3.3) \quad \beta_d = \underset{\{\beta: \beta \in L^2(S_{\tilde{w}})\}}{\operatorname{argmin}} \mathbb{E}\left\{\frac{y}{x_s^r} - \beta(\tilde{w})\right\}^2.$$

Observe that we can also characterize $x_s^r \beta_d(\tilde{w})$ as the orthogonal projection of y onto \mathcal{G}_r using the “weighted” inner product $\langle u, v \rangle_{weighted} = \mathbb{E}\{uvx_s^{-2r}\}$. But since this projection is not orthogonal w.r.t the “usual” L^2 inner product, we prefer to describe $x_s^r \beta_d(\tilde{x})$ as the “direct” or “numeraire” (population level) specification of $f(\tilde{x})$.

The preceding discussion shows that in order to estimate f it suffices to estimate β_p and β_d . In fact, since a finite amount of data can at best allow us to estimate the value taken by a function at a certain point, we consider estimating the value of f at some fixed point in $S_{\tilde{x}}^b$. So let \tilde{x}_0 be an arbitrary fixed point in $\operatorname{int}(S_{\tilde{x}}^b)$ and let $\tilde{w}_0 = H_0(\tilde{x}_0)$ denote its image under H_0 . Since the map $\tilde{x} \mapsto (\tilde{w}, x_s)$ is one to one and continuous on $S_{\tilde{x}}$, it is straightforward to verify that $\tilde{w}_0 \in \operatorname{int}(S_{\tilde{w}})$. Let us now see how we can estimate $\beta_p(\tilde{w}_0)$ and $\beta_d(\tilde{w}_0)$.

Let β be a function in $L^2(S_{\tilde{w}})$ which is sufficiently smooth. Think of β as being a generic symbol for β_p or β_d . Taylor expanding $\beta(\tilde{w})$ around \tilde{w}_0 and neglecting all higher order remainder terms, we can write

$$(3.4) \quad \beta(\tilde{w}) \doteq \beta(\tilde{w}_0) + \nabla \beta(\tilde{w}_0)'(\tilde{w} - \tilde{w}_0).$$

The unknown coefficients $\{\beta(\tilde{w}_0), \nabla \beta(\tilde{w}_0)\}$ can be estimated by doing least squares on sample analogs of (3.2) and (3.3) provided we can maintain the quality of the linear approximation in (3.4). This can be achieved by employing the usual device of “local weighting” to ensure that, when estimating $\{\beta(\tilde{w}_0), \nabla \beta(\tilde{w}_0)\}$, the \tilde{w} 's close to \tilde{w}_0 are given more weight than the observations farther away from \tilde{w}_0 .

Following this strategy we can estimate $\beta_p(\tilde{w}_0)$ and $\beta_d(\tilde{w}_0)$ by solving the following optimization problems, where, for convenience, we use the notation $\tilde{b} = (b_0, b_1, \dots, b_{s-1})$ and $\tilde{b}_* = (b_1, \dots, b_{s-1})$:

$$\{\hat{\beta}_p, \widehat{\nabla\beta}_p\}(\tilde{w}_0) = \operatorname{argmin}_{\tilde{b} \in \mathbb{R}^s} \sum_{j=1}^n \{y_j - x_{s,j}^r [b_0 + \tilde{b}'_*(\tilde{w}_j - \tilde{w}_0)]\}^2 \mathcal{K}\left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right),$$

$$\{\hat{\beta}_d, \widehat{\nabla\beta}_d\}(\tilde{w}_0) = \operatorname{argmin}_{\tilde{b} \in \mathbb{R}^s} \sum_{j=1}^n \left\{ \frac{y_j}{x_{s,j}^r} - b_0 - \tilde{b}'_*(\tilde{w}_j - \tilde{w}_0) \right\}^2 \mathcal{K}\left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right).$$

The kernel \mathcal{K} (defined on \mathbb{R}^{s-1}) and the bandwidth a_n used above satisfy the following conditions.

Assumption 3.1. $\mathcal{K}(\tilde{w}) = \prod_{i=1}^{s-1} \kappa(w_i)$ where $\kappa : \mathbb{R} \rightarrow \mathbb{R}$ is a nonnegative, symmetric, continuous function which vanishes outside $[-1, 1]$ and satisfies $\int_{-1}^1 \kappa(u) du = 1$. $\mathfrak{R}_\kappa = \int_{-1}^1 \kappa^2(u) du$ denotes the “roughness” of κ , $\mu_\kappa^2 = \int_{-1}^1 u^2 \kappa(u) du$, and $S_\mathcal{K} = [-1, 1]^{s-1}$. \square

Assumption 3.2. The bandwidth a_n is a sequence of positive numbers such that $a_n \downarrow 0$, $na_n^{s-1} \rightarrow \infty$, and $na_n^{s+3} \rightarrow \lambda \in [0, \infty)$. \square

While $\hat{\beta}_p(\tilde{w}_0)$ (resp. $\hat{\beta}_d(\tilde{w}_0)$) denotes the “projection based” (resp. “direct”) local linear estimate of $f(\tilde{w}_0, 1)$, $\widehat{\nabla\beta}_p(\tilde{w}_0)$ (resp. $\widehat{\nabla\beta}_d(\tilde{w}_0)$) is used to denote the “projection based” (resp. “direct”) local linear estimate of its gradient $\nabla f(\tilde{w}_0, 1)$. Then $\hat{f}_{\text{proj}}(\tilde{x}_0) = x_{s,0}^r \hat{\beta}_p(\tilde{w}_0)$ and $\hat{f}_{\text{dir}}(\tilde{x}_0) = x_{s,0}^r \hat{\beta}_d(\tilde{w}_0)$ denote the “projection based” and “direct” local linear estimates of $f(\tilde{x}_0)$. Note that by construction \hat{f}_{proj} and \hat{f}_{dir} are homogeneous of degree r ; i.e. we have obtained “homogeneity constrained” nonparametric estimates of f . Another nice feature of using the local linear approach is that both $f(\tilde{w}_0, 1)$ and its partial derivatives can be obtained simultaneously. This comes in handy when one wants to calculate marginal effects or elasticities. Furthermore, solving these optimization problems is straightforward since they can be expressed in a weighted least squares framework. See Lemma A.1 for details.

Remark 3.2. (i) Local linear estimates of conditional mean functions, without any homogeneity restriction, have been extensively studied, especially in the statistics literature. See, for instance, Fan (1992), Ruppert and Wand (1994) and the references therein. If instead of a first order approximation in (3.4) we had taken a m^{th} order Taylor expansion of $\beta(\tilde{w})$ around \tilde{w}_0 , where $m > 1$, we would obtain m^{th} order local polynomial estimates of β_p and β_d . In our case such higher order approximations are unnecessary since a linear approximation suffices to compare the asymptotic mse of $\hat{\beta}_p(\tilde{w}_0)$ and $\hat{\beta}_d(\tilde{w}_0)$. Of course, we could also have obtained locally constant (better known as Nadaraya-Watson) estimates of $\beta_p(\tilde{w}_0)$ and $\beta_d(\tilde{w}_0)$ by considering the “zeroth” order approximation $\beta(\tilde{w}) \doteq \beta(\tilde{w}_0)$ in (3.4). In fact it is easy to

show that the “projection based” and “direct” Nadaraya-Watson estimates of $f(\tilde{x}_0)$ are given by

$$\check{f}_{\text{proj}}(\tilde{x}_0) = x_{s,0}^r \frac{\sum_{j=1}^n y_j x_{s,j}^r \mathcal{K}\left(\frac{\tilde{w}_0 - \tilde{w}_j}{a_n}\right)}{\sum_{j=1}^n x_{s,j}^{2r} \mathcal{K}\left(\frac{\tilde{w}_0 - \tilde{w}_j}{a_n}\right)}, \quad \check{f}_{\text{dir}}(\tilde{x}_0) = x_{s,0}^r \frac{\sum_{j=1}^n \frac{y_j}{x_{s,j}^r} \mathcal{K}\left(\frac{\tilde{w}_0 - \tilde{w}_j}{a_n}\right)}{\sum_{j=1}^n \mathcal{K}\left(\frac{\tilde{w}_0 - \tilde{w}_j}{a_n}\right)}.$$

The reason we prefer working with local linear estimators, rather than the locally constant estimators, is that the asymptotic bias terms for the former are simpler (and thus much easier) to handle analytically. In particular, as shown in Lemmas A.1 and A.2, they do not have any terms involving the first derivatives of $f(\tilde{w}, 1)$:

$$\begin{aligned} \text{bias}\{\hat{f}_{\text{proj}}(\tilde{x}_0)\} &= 0.5\lambda^{1/2} \mu_\kappa^2 x_{s,0}^r \text{tr} \left\{ \frac{\partial^2 f(\tilde{w}_0, 1)}{\partial \tilde{w} \partial \tilde{w}'} \right\}, \\ \text{bias}\{\hat{f}_{\text{dir}}(\tilde{x}_0)\} &= 0.5\lambda^{1/2} \mu_\kappa^2 x_{s,0}^r \text{tr} \left\{ \frac{\partial^2 f(\tilde{w}_0, 1)}{\partial \tilde{w} \partial \tilde{w}'} \right\}. \end{aligned}$$

Hence, when local linear estimators are used

$$\text{bias}\{\hat{f}_{\text{dir}}(\tilde{x}_0)\} = \text{bias}\{\hat{f}_{\text{proj}}(\tilde{x}_0)\},$$

and so comparing the asymptotic *mse* of $\hat{f}_{\text{proj}}(\tilde{x}_0)$ and $\hat{f}_{\text{dir}}(\tilde{x}_0)$ reduces to comparing their asymptotic variances. In contrast, in Appendix B it is shown that for the Nadaraya-Watson estimators of $f(\tilde{x}_0)$,

$$\begin{aligned} \text{bias}\{\check{f}_{\text{proj}}(\tilde{x}_0)\} &\doteq 0.5a_n^2 \mu_\kappa^2 x_{s,0}^r \text{tr} \left\{ \frac{\partial^2 f(\tilde{w}_0, 1)}{\partial \tilde{w} \partial \tilde{w}'} + \frac{2}{h(\tilde{w}_0)} \times \right. \\ &\quad \left[\frac{\partial f(\tilde{w}_0, 1)}{\partial \tilde{w}} \frac{\partial h(\tilde{w}_0)}{\partial \tilde{w}'} + \frac{\partial f(\tilde{w}_0, 1)}{\partial \tilde{w}} \frac{\partial \mathbb{E}(x_s^{2r} | \tilde{w}_0)}{\partial \tilde{w}'} \frac{h(\tilde{w}_0)}{\mathbb{E}(x_s^{2r} | \tilde{w}_0)} \right] \Big\}, \\ \text{bias}\{\check{f}_{\text{dir}}(\tilde{x}_0)\} &\doteq 0.5a_n^2 \mu_\kappa^2 x_{s,0}^r \text{tr} \left\{ \frac{\partial^2 f(\tilde{w}_0, 1)}{\partial \tilde{w} \partial \tilde{w}'} + \frac{2}{h(\tilde{w}_0)} \times \right. \\ &\quad \left. \frac{\partial f(\tilde{w}_0, 1)}{\partial \tilde{w}} \frac{\partial h(\tilde{w}_0)}{\partial \tilde{w}'} \right\}. \end{aligned}$$

Although in Appendix B we also show that $\text{var}\{\check{f}_{\text{proj}}(\tilde{x}_0)\} = \text{var}\{\hat{f}_{\text{proj}}(\tilde{x}_0)\}$ and $\text{var}\{\check{f}_{\text{dir}}(\tilde{x}_0)\} = \text{var}\{\hat{f}_{\text{dir}}(\tilde{x}_0)\}$, the squared bias of $\check{f}_{\text{proj}}(\tilde{x}_0)$ and $\check{f}_{\text{dir}}(\tilde{x}_0)$ cannot be ranked. Therefore, if we use Nadaraya-Watson estimators we cannot analytically compare a “projection based” estimator with the “direct” estimator in terms of asymptotic *mse*.

(ii) The reader should also be aware that by confining attention to $\hat{f}_{\text{proj}}(\tilde{x}_0)$ and $\hat{f}_{\text{dir}}(\tilde{x}_0)$ we are ignoring the possibility that we could perhaps do better by (say) looking at a linear combination of the two. Indeed such a possibility cannot be ruled out unless we can show that one of these estimators is “efficient” (in some sense). However, the task of constructing “efficient” nonparametric estimates of $f(\tilde{x}_0)$ is beyond the scope of this paper. In fact we are unaware of any “nonparametric efficiency bound” type result in the literature which would allow us to determine the efficiency of a proposed nonparametric estimator of f under a homogeneity restriction.

(iii) Let $\hat{f}(\tilde{x}_0)$ denote the “usual” local linear estimator of $f(\tilde{x}_0)$ in the regression model $y = f(\tilde{x}) + \varepsilon$; i.e $\hat{f}(\tilde{x}_0)$ solves

$$\{\hat{f}(\tilde{x}_0), \widehat{\nabla} f(\tilde{x}_0)\} = \operatorname{argmin}_{\tilde{b} \in \mathbb{R}^s} \sum_{j=1}^n \{y_j - b_0 - \tilde{b}'(\tilde{x}_j - \tilde{x}_0)\}^2 \mathcal{H}\left(\frac{\tilde{x}_j - \tilde{x}_0}{b_n}\right),$$

where \mathcal{H} is an appropriate kernel on \mathbb{R}^s , and the bandwidth b_n is a sequence of positive numbers such that $b_n \downarrow 0$ and $nb_n^s \rightarrow \infty$. Note that when f is homogeneous of degree r , \hat{f} will converge at a slower rate to f than \hat{f}_{proj} or \hat{f}_{dir} . Imposing homogeneity on estimators of f reduces the dimension of the regressor space by one, and leads to estimators with improved rates of convergence. In the simulations we will compare the finite sample behavior of \hat{f}_{proj} and \hat{f}_{dir} with \hat{f} . \square

4. COMPARING $\hat{f}_{\text{proj}}(\tilde{x}_0)$ AND $\hat{f}_{\text{dir}}(\tilde{x}_0)$

Following the results of Lemma A.1 and A.2, it is easy to see that the asymptotic variances of $\hat{f}_{\text{proj}}(\tilde{x}_0)$ and $\hat{f}_{\text{dir}}(\tilde{x}_0)$ are

$$\begin{aligned} \operatorname{var}\{\hat{f}_{\text{proj}}(\tilde{x}_0)\} &= \frac{x_{s,0}^{2r} \mathbb{E}(x_s^{2r} \varepsilon^2 | \tilde{w}_0) \mathfrak{R}_\kappa^{s-1}}{\mathbb{E}^2(x_s^{2r} | \tilde{w}_0) h(\tilde{w}_0)}, \\ \operatorname{var}\{\hat{f}_{\text{dir}}(\tilde{x}_0)\} &= \frac{x_{s,0}^{2r} \mathbb{E}(x_s^{-2r} \varepsilon^2 | \tilde{w}_0) \mathfrak{R}_\kappa^{s-1}}{h(\tilde{w}_0)}. \end{aligned}$$

To simplify the form of these variances observe that the transformation $\tilde{x} \mapsto (\tilde{w}, x_s)$ is one to one and apply iterated expectations. This yields

$$\mathbb{E}(x_s^{2r} \varepsilon^2 | \tilde{w}) = \mathbb{E}(x_s^{2r} \sigma^2(\tilde{x}) | \tilde{w}), \quad \mathbb{E}(x_s^{-2r} \varepsilon^2 | \tilde{w}) = \mathbb{E}(x_s^{-2r} \sigma^2(\tilde{x}) | \tilde{w}),$$

where the conditional variance function $\sigma^2(\tilde{x}) = \operatorname{var}(\varepsilon | \tilde{x})$. It does not seem possible (at least to us) to compare the two variances if $\sigma^2(\tilde{x})$ is completely unknown. However, it is possible to obtain some useful insights about the asymptotic variance of $\hat{f}_{\text{proj}}(\tilde{x}_0)$ and $\hat{f}_{\text{dir}}(\tilde{x}_0)$ if $\sigma^2(\tilde{x})$ satisfies the following assumption.

Assumption 4.1. *Assume that either*

- (i) $\sigma^2(\tilde{x}) = \psi^2(\tilde{w})$ for some unknown ψ ; i.e. the error terms are conditionally heteroscedastic such that $\sigma(\tilde{x})$ is homogeneous of degree zero in the covariates, or
- (ii) $\sigma^2(\tilde{x}) = x_s^{2r} \psi^2(\tilde{w})$ for some unknown ψ ; i.e. the error terms are conditionally heteroscedastic such that $\sigma(\tilde{x})$ is homogeneous of degree $r \neq 0$ in the covariates. \square

Notice that Assumption 4.1(i) is automatically satisfied if the error term ε is assumed to be homoscedastic. As an example of a model where homoscedasticity of ε is compatible with linear homogeneity of f , consider the following simple setup.

Example 4.1. Let y be the observed profit, f the unobserved profit function of a competitive firm, and \tilde{x} the vector of observed input and factor prices. Assuming that the prices are measured without error but there is measurement error in the observed profit, we can write the model to be estimated as $y = f(\tilde{x}) + \varepsilon$. Since f is a profit function it is homogeneous of degree one, and as ε is treated as pure measurement error we can assume that it is homoscedastic. \square

Although homoscedasticity of additive errors is often a convenient statistical assumption, it is sometimes hard to justify from a structural point of view. In many cases Assumption 4.1(ii) may be more plausible. As an example of a situation where both $\sigma(\tilde{x})$ and $f(\tilde{x})$ are homogeneous of degree 1, consider the following model which is motivated by the discussion in McFadden (1984, Page 1406).

Example 4.2. Let \tilde{x} be the vector of observed input and factor prices and $f^*(\tilde{x}; u)$ the unobserved profit function of a competitive firm. u is a firm specific random parameter which is distributed independently of \tilde{x} . It is unobserved by the researcher but is known to the firm itself. For instance, u could represent variables which are unobserved by the economist but are used by the firm when making production decisions. Since f^* is a profit function, we assume that for each $u \in U$ the map $\tilde{x} \mapsto f^*(\tilde{x}; u)$ is linearly homogeneous, monotone, and convex in the prices. Let $y = f^*(\tilde{x}; u)$ denote the maximum observable profit. Then using the fact that u is independent of \tilde{x} , we can write $y = f(\tilde{x}) + \varepsilon$ where $f(\tilde{x}) = \int_U f^*(\tilde{x}; u) dF(u)$, and

$$\varepsilon = f^*(\tilde{x}; u) - \int_U f^*(\tilde{x}; u) dF(u|\tilde{x}) = f^*(\tilde{x}; u) - \int_U f^*(\tilde{x}; u) dF(u).$$

Note that $\tilde{x} \mapsto f(\tilde{x})$ has all the properties of a profit function; i.e. in particular it is homogeneous of degree one. Moreover, we can verify that $\mathbb{E}(\varepsilon|\tilde{x}) = 0$, and $\sigma(\tilde{x})$ is homogeneous of degree 1. \square

Now it is easy to see that

$$\begin{aligned} \text{Assumption 4.1(i)} &\implies \begin{cases} \mathbb{E}\{x_s^{-2r}\varepsilon^2|\tilde{w}\} = \psi^2(\tilde{w}) \mathbb{E}\{x_s^{-2r}|\tilde{w}\} \\ \mathbb{E}\{x_s^{2r}\varepsilon^2|\tilde{w}\} = \psi^2(\tilde{w}) \mathbb{E}\{x_s^{2r}|\tilde{w}\}, \end{cases} \\ \text{Assumption 4.1(ii)} &\implies \begin{cases} \mathbb{E}\{x_s^{-2r}\varepsilon^2|\tilde{w}\} = \psi^2(\tilde{w}) \\ \mathbb{E}\{x_s^{2r}\varepsilon^2|\tilde{w}\} = \psi^2(\tilde{w}) \mathbb{E}\{x_s^{4r}|\tilde{w}\}. \end{cases} \end{aligned}$$

Hence the expressions for the asymptotic variances simplify to

$$\begin{aligned} \text{Assumption 4.1(i)} &\implies \begin{cases} \text{var}\{\hat{f}_{\text{dir}}(\tilde{x}_0)\} = \frac{\psi^2(\tilde{w}_0) \mathfrak{R}_\kappa^{s-1} x_{s,0}^{2r} \mathbb{E}(x_s^{-2r}|\tilde{w}_0)}{h(\tilde{w}_0)} \\ \text{var}\{\hat{f}_{\text{proj}}(\tilde{x}_0)\} = \frac{\psi^2(\tilde{w}_0) \mathfrak{R}_\kappa^{s-1} x_{s,0}^{2r}}{\mathbb{E}(x_s^{2r}|\tilde{w}_0)h(\tilde{w}_0)}, \end{cases} \\ \text{Assumption 4.1(ii)} &\implies \begin{cases} \text{var}\{\hat{f}_{\text{dir}}(\tilde{x}_0)\} = \frac{\psi^2(\tilde{w}_0) \mathfrak{R}_\kappa^{s-1} x_{s,0}^{2r}}{h(\tilde{w}_0)} \\ \text{var}\{\hat{f}_{\text{proj}}(\tilde{x}_0)\} = \frac{\psi^2(\tilde{w}_0) \mathfrak{R}_\kappa^{s-1} x_{s,0}^{2r} \mathbb{E}(x_s^{4r}|\tilde{w}_0)}{\mathbb{E}^2(x_s^{2r}|\tilde{w}_0)h(\tilde{w}_0)}. \end{cases} \end{aligned}$$

Since by the Cauchy-Schwarz inequality

$$\mathbb{E}(x_s^{-2r}|\tilde{w})\mathbb{E}(x_s^{2r}|\tilde{w}) \geq \mathbb{E}^2(x_s^{-r}x_s^r|\tilde{w}) = 1 \text{ and } \mathbb{E}(x_s^{4r}|\tilde{w}) \geq \mathbb{E}^2(x_s^{2r}|\tilde{w}),$$

we get that

$$\text{var}\{\hat{f}_{\text{proj}}(\tilde{x}_0)\} \leq \text{var}\{\hat{f}_{\text{dir}}(\tilde{x}_0)\},$$

under Assumption 4.1(i), and

$$\text{var}\{\hat{f}_{\text{dir}}(\tilde{x}_0)\} \leq \text{var}\{\hat{f}_{\text{proj}}(\tilde{x}_0)\}$$

under Assumption 4.1(ii). As we had already shown that the asymptotic bias terms for the two estimators were the same, we get that $mse\{\hat{f}_{\text{proj}}(\tilde{x}_0)\} \leq mse\{\hat{f}_{\text{dir}}(\tilde{x}_0)\}$ under Assumption 4.1(i), and $mse\{\hat{f}_{\text{dir}}(\tilde{x}_0)\} \leq mse\{\hat{f}_{\text{proj}}(\tilde{x}_0)\}$ under Assumption 4.1(ii). Therefore, as expected, there is no general ranking for the estimators in terms of asymptotic mse . Hence the choice of which estimate to use is not obvious but depends upon the nature of the heteroscedasticity of the error term.

5. SIMULATION

A small simulation experiment was performed to study the finite sample properties of the proposed estimators. Code was written in GAUSS and we restricted our attention to the case $s = 2$. n observations on (y, x_1, x_2) were generated from $y = f(x_1, x_2) + \sigma(x_1, x_2)\varepsilon$, where $x_1, x_2 \stackrel{d}{=} \text{UIID}[1, 2]$ and ε was chosen independently of (x_1, x_2) . A Gaussian kernel was used to estimate \hat{f}_{proj} , \hat{f}_{dir} and \hat{f} . For the first two estimators the bandwidth used was $cn^{-1/5}$, while for \hat{f} the bandwidth used was $cn^{-1/6}$. Three different choices of c were considered: $c \in \{0.5, 1, 2\}$. As seen in Tables 1 and 2, the results do not seem to be very sensitive to the choice of bandwidth. Two particular specifications for f and ε were selected.

$$\text{(Model 1)} \quad f_1(x_1, x_2) = 10\sqrt{x_1x_2}, \quad \varepsilon_1 \stackrel{d}{=} \text{N}(0, 0.75)$$

$$\text{(Model 2)} \quad f_2(x_1, x_2) = 10(x_1^{0.5} + x_2^{0.5})^2, \quad \varepsilon_2 \stackrel{d}{=} \text{N}(0, 1).$$

f_1 and f_2 represent a Cobb-Douglas and CES specification respectively. Note that both f_1 and f_2 are homogeneous of degree one. The model parameters were chosen so that $\text{var}\{f(x_1, x_2)\}/\text{var}\{y\}$, which can be thought of as a measure of the $S/(S + N)$ ratio, for each model is around 0.8. Two simple values of $\sigma(x_1, x_2)$ were chosen to satisfy Assumption 4.1: $\sigma(x_1, x_2) = 1$ which satisfies 4.1(i), and $\sigma(x_1, x_2) = x_2$ which satisfies 4.1(ii). Each function was estimated at a 10×10 uniform grid in $[1, 2] \times [1, 2]$ and the mse calculated at each grid point in 1000 replications.

Recall that our asymptotic results are about the pointwise behavior of the mse . But since the reader may find point by point comparison of mse on the 10×10 grid a tedious task, we present the average (over 100 grid points) mse in Tables 1 and 2.

TABLE 1. Average MSE over grid for Model 1 (Cobb-Douglas).

n	Average MSE	$\sigma(x_1, x_2) = 1$			$\sigma(x_1, x_2) = x_2$		
		$c = 0.5$	$c = 1$	$c = 2$	$c = 0.5$	$c = 1$	$c = 2$
50	\hat{f}_{proj}	0.046	0.039	0.045	0.101	0.083	0.087
	\hat{f}_{dir}	0.049	0.042	0.051	0.094	0.078	0.082
	\hat{f}	0.074	0.054	0.060	0.164	0.116	0.120
100	\hat{f}_{proj}	0.023	0.023	0.029	0.050	0.044	0.048
	\hat{f}_{dir}	0.025	0.025	0.032	0.047	0.042	0.047
	\hat{f}	0.038	0.031	0.036	0.085	0.062	0.063

TABLE 2. Average MSE over 100 grid points for Model 2 (CES).

n	Average MSE	$\sigma(x_1, x_2) = 1$			$\sigma(x_1, x_2) = x_2$		
		$c = 0.5$	$c = 1$	$c = 2$	$c = 0.5$	$c = 1$	$c = 2$
50	\hat{f}_{proj}	0.122	0.120	0.139	0.275	0.230	0.266
	\hat{f}_{dir}	0.128	0.132	0.151	0.258	0.221	0.251
	\hat{f}	0.196	0.162	0.175	0.448	0.324	0.352
100	\hat{f}_{proj}	0.066	0.067	0.097	0.146	0.132	0.148
	\hat{f}_{dir}	0.069	0.074	0.107	0.136	0.128	0.150
	\hat{f}	0.105	0.090	0.118	0.247	0.179	0.188

As seen in the tables, the average mse is ranked according to our asymptotic results. Moreover, except in one case², this ranking does not change when c is varied. When ε is homoscedastic, \hat{f}_{proj} dominates, although there does not seem to be a dramatic difference between \hat{f}_{proj} and \hat{f}_{dir} in terms of average mse . When the conditional variance of ε is homogeneous of degree 1, \hat{f}_{dir} has the smallest average mse although once again the difference between \hat{f}_{proj} and \hat{f}_{dir} is not very substantial. However, in each case, the homogeneity constrained estimators clearly out perform \hat{f} , the usual local linear estimate of f . Therefore, when f is indeed homogeneous, using a homogeneity constrained estimator seems sensible. Of course, the reader must keep in mind the usual caveat about any simulations results; namely, that they are limited in nature and may vary if the underlying model parameters are changed.

²For Model 2 when $n = 100$ and $c = 2$. However, a decomposition of the average mse for this case revealed that while the average variance for \hat{f}_{dir} was smaller than the average variance of \hat{f}_{proj} , the average squared bias for the former was bigger than the average squared bias for the latter. Therefore, while the ranking w.r.t the variances is preserved according to our theory, the average bias differs for the two models. Hence, in this case the higher average bias of \hat{f}_{dir} caused the ranking to change.

6. CONCLUSION

In this paper we nonparametrically estimate a homogeneous of degree r conditional mean function (f) using local linear estimators. We compare a “projection based” estimator with a more conventional “direct” estimator. Based on our asymptotic results we recommend the following guidelines when estimating f in practice:

- (i) When f is homogeneous of degree r , use a “homogeneity constrained” estimator as opposed to some usual “unrestricted” nonparametric estimator of f . The dimension reduction due to homogeneity allows the constrained estimators to possess faster rates of convergence than the unrestricted estimator.
- (ii) Use the “projection based” approach if $\sigma(\tilde{x})$ is homogeneous of degree zero. This includes the case when ε is homoscedastic.
- (iii) Use the “direct” approach if $\sigma(\tilde{x})$ is homogeneous of degree $r \neq 0$.

Results of a small simulation experiment support these recommendations although there does not seem to be a big difference in the average mse for the two approaches in (ii) and (iii), at least for the models used in our simulation.

Our theoretical results do not offer any guidance if no information is available about the conditional variance function. In such a case it may be good empirical practice to report estimates of f using the “projection based,” “direct,” as well as the “unrestricted” local linear estimate of f . If there is wide divergence between the reported results it may be an indication that the homogeneity restriction on f is perhaps misspecified. Based on the degree of divergence a formal test of this misspecification can be constructed following the approach of Härdle and Mammen (1993), although we do not pursue this issue in the current paper.

Finally, the reader should bear in mind that in this paper we have limited our investigations to the case when f is homogeneous of degree r . But, as mentioned earlier, in microeconomic theory homogeneity of functional forms is often accompanied with other shape restrictions such as monotonicity and concavity (or convexity). An interesting topic for future research is to find new ways of nonparametrically imposing these additional shape restrictions on functional forms and determine the statistical properties of such shape restricted estimators. \square

APPENDIX A. TECHNICAL DETAILS: LOCAL LINEAR ESTIMATORS

Lemma A.1. *Let assumptions 2.1, 3.1, and 3.2 hold. Then*

$$\left[\begin{array}{c} \sqrt{na_n^{s-1}}\{\hat{\beta}_p(\tilde{w}_0) - f(\tilde{w}_0, 1) - Bias_1\} \\ \sqrt{na_n^{s+1}}\{\widehat{\nabla}\beta_p(\tilde{w}_0) - \nabla f(\tilde{w}_0, 1) - Bias_2\} \end{array} \right] \xrightarrow{d} N(0, \Sigma),$$

where $\begin{bmatrix} Bias_1 \\ Bias_2 \end{bmatrix} = \begin{bmatrix} \frac{a_n^2}{2} \mu_\kappa^2 tr\{\nabla^2 f(\tilde{w}_0, 1)\} \\ \frac{a_n}{2\mu_\kappa^2} \int_{S_\kappa} \tilde{u} \mathcal{K}(\tilde{u}) \tilde{u}' \nabla^2 f(\tilde{w}_0, 1) \tilde{u} d\tilde{u} \end{bmatrix}$, and the asymptotic variance covariance matrix

$$\Sigma = \begin{bmatrix} \frac{\mathbb{E}(x_s^{2r} \varepsilon^2 | \tilde{w}_0) \mathfrak{A}_\kappa^{s-1}}{\mathbb{E}^2(x_s^{2r} | \tilde{w}_0) h(\tilde{w}_0)} & 0'_{s-1} \\ 0_{s-1} & \frac{\mathbb{E}(x_s^{2r} \varepsilon^2 | \tilde{w}_0) \int_{S_\kappa} \tilde{u} \tilde{u}' \mathcal{K}^2(\tilde{u}) d\tilde{u}}{\mathbb{E}^2(x_s^{2r} | \tilde{w}_0) h(\tilde{w}_0) \mu_\kappa^4} \end{bmatrix}.$$

Proof. Throughout this proof let $Q_n = diag_{s \times s} [1, a_n, \dots, a_n]$ and let Ω_n denote the diagonal matrix $diag_{n \times n} [\mathcal{K}(\frac{\tilde{w}_1 - \tilde{w}_0}{a_n}), \dots, \mathcal{K}(\frac{\tilde{w}_n - \tilde{w}_0}{a_n})]$. Furthermore, for notational convenience, we also let

$$\hat{\theta}(\tilde{w}_0) = \begin{bmatrix} \hat{\theta}_0(\tilde{w}_0) \\ \hat{\theta}_1(\tilde{w}_0) \\ \vdots \\ \hat{\theta}_{s-1}(\tilde{w}_0) \end{bmatrix}_{s \times 1} = \begin{bmatrix} \hat{\beta}_p(\tilde{w}_0) \\ \nabla \hat{\beta}_p(\tilde{w}_0) \end{bmatrix}, \quad Z_n = \begin{bmatrix} x_{s,1}^r & x_{s,1}^r (\frac{\tilde{w}'_1 - \tilde{w}'_0}{a_n}) \\ \vdots & \vdots \\ x_{s,n}^r & x_{s,n}^r (\frac{\tilde{w}'_n - \tilde{w}'_0}{a_n}) \end{bmatrix}_{n \times s}.$$

Using this notation it is straightforward to see that $\hat{\theta}(\tilde{w}_0)$ is the solution to the following weighted least squares problem:

$$\min_{\tilde{b}} (\tilde{y} - Z_n Q_n \tilde{b})' \Omega_n (\tilde{y} - Z_n Q_n \tilde{b}).$$

As is well known, the solution to this problem is given by

$$\hat{\theta}(\tilde{w}_0) = Q_n^{-1} (Z_n' \Omega_n Z_n)^{-1} Z_n' \Omega_n \tilde{y}.$$

Let us write $\hat{\theta}(\tilde{w}_0) = Q_n^{-1} S_n^{-1} \tilde{t}_n$, where

$$S_n = \frac{1}{n a_n^{s-1}} Z_n' \Omega_n Z_n \quad \text{and} \quad \tilde{t}_n = \frac{1}{n a_n^{s-1}} Z_n' \Omega_n \tilde{y}.$$

Straightforward calculations show that we can write S_n as the partitioned matrix $S_n = \begin{bmatrix} s_{00} & s'_{01} \\ s_{01} & s_{11} \end{bmatrix}$, where $s_{00} = \frac{1}{n a_n^{s-1}} \sum_{j=1}^n x_{s,j}^{2r} \mathcal{K}(\frac{\tilde{w}_j - \tilde{w}_0}{a_n})$,

$$s_{01} = \frac{1}{n a_n^{s-1}} \sum_{j=1}^n x_{s,j}^{2r} (\frac{\tilde{w}_j - \tilde{w}_0}{a_n}) \mathcal{K}(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}),$$

$$s_{11} = \frac{1}{n a_n^{s-1}} \sum_{j=1}^n x_{s,j}^{2r} (\frac{\tilde{w}_j - \tilde{w}_0}{a_n}) (\frac{\tilde{w}'_j - \tilde{w}'_0}{a_n}) \mathcal{K}(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}).$$

Similarly, we write $\tilde{t}_n = \begin{bmatrix} t_0 \\ t_1 \end{bmatrix}$ where $t_0 = \frac{1}{n a_n^{s-1}} \sum_{j=1}^n y_j x_{s,j}^r \mathcal{K}(\frac{\tilde{w}_j - \tilde{w}_0}{a_n})$, and $t_1 = \frac{1}{n a_n^{s-1}} \sum_{j=1}^n y_j x_{s,j}^r (\frac{\tilde{w}_j - \tilde{w}_0}{a_n}) \mathcal{K}(\frac{\tilde{w}_j - \tilde{w}_0}{a_n})$. But since $y_j = x_{s,j}^r f(\tilde{w}_j, 1) + \varepsilon_j$

we can express $\tilde{t}_n = \tilde{\tau}_n + \tilde{t}_n^*$, where

$$\tilde{\tau}_n = \begin{bmatrix} \tau_0 \\ \tau_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{na_n^{s-1}} \sum_{j=1}^n \varepsilon_j x_{s,j}^r \mathcal{K}\left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right) \\ \frac{1}{na_n^{s-1}} \sum_{j=1}^n \varepsilon_j x_{s,j}^r \left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right) \mathcal{K}\left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right) \end{bmatrix},$$

$$\tilde{t}_n^* = \begin{bmatrix} t_0^* \\ t_1^* \end{bmatrix} = \begin{bmatrix} \frac{1}{na_n^{s-1}} \sum_{j=1}^n x_{s,j}^{2r} f(\tilde{w}_j, 1) \mathcal{K}\left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right) \\ \frac{1}{na_n^{s-1}} \sum_{j=1}^n x_{s,j}^{2r} f(\tilde{w}_j, 1) \left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right) \mathcal{K}\left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right) \end{bmatrix}.$$

Hence, letting $\theta(\tilde{w}_0) = \begin{bmatrix} f(\tilde{w}_0, 1) \\ \nabla f(\tilde{w}_0, 1) \end{bmatrix}$ denote the probability limit of $\hat{\theta}(\tilde{w}_0)$, we have that

$$(A.1) \quad \hat{\theta}(\tilde{w}_0) - \theta(\tilde{w}_0) = Q_n^{-1} S_n^{-1} \tilde{\tau}_n + Q_n^{-1} S_n^{-1} \tilde{t}_n^* - \theta(\tilde{w}_0).$$

Let us first look at $Q_n^{-1} S_n^{-1} \tilde{t}_n^* - \theta(\tilde{w}_0)$. For all \tilde{w}_j in a a_n -neighborhood of \tilde{w}_0 , Taylor expand $f(\tilde{w}_j, 1)$ around $f(\tilde{w}_0, 1)$ to get

$$f(\tilde{w}_j, 1) = f(\tilde{w}_0, 1) + a_n \left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right)' \nabla f(\tilde{w}_0, 1) \\ + \frac{a_n^2}{2} \left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right)' \nabla^2 f(\tilde{w}_0, 1) \left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right) + o(a_n^2).$$

But this implies that \tilde{t}_n^* reduces to

$$\tilde{t}_n^* = S_n Q_n \theta(\tilde{w}_0) + \frac{a_n^2}{2} B_n + \begin{bmatrix} o(a_n^2) \\ \vdots \\ o(a_n^2) \end{bmatrix}_{s \times 1},$$

where

$$B_n = \begin{bmatrix} \frac{1}{na_n^{s-1}} \sum_{j=1}^n x_{s,j}^{2r} \left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right)' \nabla^2 f(\tilde{w}_0, 1) \left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right) \mathcal{K}\left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right) \\ \frac{1}{na_n^{s-1}} \sum_{j=1}^n x_{s,j}^{2r} \left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right)' \nabla^2 f(\tilde{w}_0, 1) \left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right) \mathcal{K}\left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right) \left(\frac{\tilde{w}_j - \tilde{w}_0}{a_n}\right) \end{bmatrix}.$$

Since it is easy to verify that

$$S_n \xrightarrow{p} \begin{bmatrix} \mathbb{E}(x_s^{2r} | \tilde{w}_0) h(\tilde{w}_0) & 0'_{s-1} \\ 0_{s-1} & \mu_\kappa^2 \mathbb{E}(x_s^{2r} | \tilde{w}_0) I_{s-1} \end{bmatrix},$$

we get that

$$Q_n^{-1} S_n^{-1} \tilde{t}_n^* - \theta(\tilde{w}_0) = \frac{a_n^2}{2} Q_n^{-1} S_n^{-1} B_n + \begin{bmatrix} o_p(a_n^2) \\ o_p(a_n) \\ \vdots \\ o_p(a_n) \end{bmatrix}_{s \times 1}.$$

Let us further simplify the right hand side of this equation. To do so first use a LLN to show that

$$B_n \xrightarrow{p} \mathbb{E}(x_s^{2r} | \tilde{w}_0) h(\tilde{w}_0) \left[\int_{S_{\mathcal{K}}} \text{vec}'(\mu_\kappa^2 I_{s-1}) \tilde{u} \mathcal{K}(\tilde{u}) \text{vec}'(\tilde{u} \tilde{u}') d\tilde{u} \right] \text{vec}(\nabla^2 f(\tilde{w}_0, 1)).$$

Next, a little matrix manipulation shows that

$$\frac{a_n^2}{2} Q_n^{-1} S_n^{-1} B_n = \begin{bmatrix} Bias_1 \\ Bias_2 \end{bmatrix} + \begin{bmatrix} o_p(a_n^2) \\ o_p(a_n) \\ \vdots \\ o_p(a_n) \end{bmatrix}_{s \times 1}.$$

Therefore, letting $Bias_n = \begin{bmatrix} Bias_1 \\ Bias_2 \end{bmatrix}$, (A.1) reduces to

$$(A.2) \quad \hat{\theta}(\tilde{w}_0) - \theta(\tilde{w}_0) - Bias_n = Q_n^{-1} S_n^{-1} \tilde{\tau}_n + \begin{bmatrix} o_p(a_n^2) \\ o_p(a_n) \\ \vdots \\ o_p(a_n) \end{bmatrix}_{s \times 1}.$$

Now we show that $S_n^{-1} \tilde{\tau}_n$ is asymptotically normal. Under Assumption 2.1(v) a straightforward application of the Lindeberg-Lyapunov CLT for triangular arrays reveals that

$$\sqrt{na_n^{s-1}} \tilde{\tau}_n \xrightarrow{d} N(0_s, \mathcal{K}^* \mathbb{E}(\varepsilon^2 x_s^{2r} | \tilde{w}_0) h(\tilde{w}_0)),$$

where

$$\mathcal{K}^* = \begin{bmatrix} \int_{S_{\mathcal{K}}} \mathcal{K}^2(\tilde{u}) d\tilde{u} & 0'_{s-1} \\ 0_{s-1} & \int_{S_{\mathcal{K}}} \tilde{u} \tilde{u}' \mathcal{K}^2(\tilde{u}) d\tilde{u} \end{bmatrix}_{s \times s}.$$

Therefore, by Slutsky, $\sqrt{na_n^{s-1}} S_n^{-1} \tilde{\tau}_n \xrightarrow{d} N(0_s, \Sigma)$. Finally, premultiplying both sides of (A.2) by $\sqrt{na_n^{s-1}} Q_n$ and using the fact that the sequence na_n^{s+3} is bounded, it is immediate that

$$\sqrt{na_n^{s-1}} Q_n \{\hat{\theta}(\tilde{w}_0) - \theta(\tilde{w}_0) - Bias_n\} \xrightarrow{d} N(0_s, \Sigma).$$

But this is the desired result. \square

Lemma A.2. *Let assumptions 2.1, 3.1, and 3.2 hold. Then*

$$\begin{bmatrix} \sqrt{na_n^{s-1}} \{\hat{\beta}_d(\tilde{w}_0) - f(\tilde{w}_0, 1) - Bias_1\} \\ \sqrt{na_n^{s+1}} \{\widehat{\nabla} \beta_d(\tilde{w}_0) - \nabla f(\tilde{w}_0, 1) - Bias_2\} \end{bmatrix} \xrightarrow{d} N(0, \Sigma),$$

where $\begin{bmatrix} Bias_1 \\ Bias_2 \end{bmatrix} = \begin{bmatrix} \frac{a_n^2}{2} \mu_{\kappa}^2 tr\{\nabla^2 f(\tilde{w}_0, 1)\} \\ \frac{a_n}{2\mu_{\kappa}^2} \int_{S_{\mathcal{K}}} \tilde{u} \mathcal{K}(\tilde{u}) \tilde{u}' \nabla^2 f(\tilde{w}_0, 1) \tilde{u} d\tilde{u} \end{bmatrix}$, and

$$\Sigma = \begin{bmatrix} \frac{\mathbb{E}(x_s^{-2r} \varepsilon^2 | \tilde{w}_0) \mathfrak{A}_{\kappa}^{s-1}}{h(\tilde{w}_0)} & 0'_{s-1} \\ 0_{s-1} & \frac{\mathbb{E}(x_s^{-2r} \varepsilon^2 | \tilde{w}_0) \int_{S_{\mathcal{K}}} \tilde{u} \tilde{u}' \mathcal{K}^2(\tilde{u}) d\tilde{u}}{h(\tilde{w}_0) \mu_{\kappa}^4} \end{bmatrix}.$$

Proof. In the proof of Lemma A.1 replace $x_{s,j}^r$ by 1, y_j by $y_j/x_{s,j}^r$, and ε_j by $\varepsilon_j/x_{s,j}^r$. \square

APPENDIX B. TECHNICAL DETAILS: KERNEL ESTIMATORS

The following results are essentially an exercise in using the delta method. For examples on the use of the delta method or “linearization” techniques in nonparametric regression see Schuster (1972) and Härdle (1989).

Lemma B.1. $var\{\check{f}_{\text{proj}}(\tilde{x}_0)\} \doteq \frac{\mathfrak{R}_\kappa^{s-1} x_{s,0}^{2r} \mathbb{E}(x_s^{2r} \varepsilon^2 | \tilde{w}_0)}{n a_n^{s-1} \mathbb{E}^2(x_s^{2r} | \tilde{w}_0) h(\tilde{w}_0)}$.

Proof. Write $\check{f}_{\text{proj}}(\tilde{x}_0) = \check{A}_n(\tilde{x}_0)/\check{B}_n(\tilde{x}_0)$ and $f(\tilde{x}_0) = A(\tilde{x}_0)/B(\tilde{x}_0)$, where

$$\begin{aligned} \check{A}_n(\tilde{x}_0) &= \frac{x_{s,0}^r}{n a_n^{s-1}} \sum_{j=1}^n y_j x_{s,j}^r \mathcal{K}\left(\frac{\tilde{w}_0 - \tilde{w}_j}{a_n}\right), & A(\tilde{x}_0) &= x_{s,0}^r \mathbb{E}(y x_s^r | \tilde{w}_0) h(\tilde{w}_0) \\ \check{B}_n(\tilde{x}_0) &= \frac{1}{n a_n^{s-1}} \sum_{j=1}^n x_{s,j}^{2r} \mathcal{K}\left(\frac{\tilde{w}_0 - \tilde{w}_j}{a_n}\right), & B(\tilde{x}_0) &= \mathbb{E}(x_s^{2r} | \tilde{w}_0) h(\tilde{w}_0). \end{aligned}$$

Then by a Taylor expansion³

$$\begin{aligned} \check{f}_{\text{proj}}(\tilde{x}_0) - \mathbb{E}\check{f}_{\text{proj}}(\tilde{x}_0) &\doteq \frac{1}{B(\tilde{x}_0)} \{\check{A}_n(\tilde{x}_0) - \mathbb{E}\check{A}_n(\tilde{x}_0)\} \\ &\quad - \frac{f(\tilde{x}_0)}{B(\tilde{x}_0)} \{\check{B}_n(\tilde{x}_0) - \mathbb{E}\check{B}_n(\tilde{x}_0)\}. \end{aligned}$$

Therefore,

$$\begin{aligned} \text{(B.1)} \quad var\{\check{f}_{\text{proj}}(\tilde{x}_0)\} &= \frac{1}{B^2(\tilde{x}_0)} var\{\check{A}_n(\tilde{x}_0)\} + \frac{f^2(\tilde{x}_0)}{B^2(\tilde{x}_0)} var\{\check{B}_n(\tilde{x}_0)\} \\ &\quad - 2 \frac{f(\tilde{x}_0)}{B^2(\tilde{x}_0)} cov\{\check{A}_n(\tilde{x}_0), \check{B}_n(\tilde{x}_0)\}. \end{aligned}$$

Recall that \tilde{w}_0 lies in the interior of $S_{\tilde{w}}$ and that the maps $\mathbb{E}(y x_s^r | \tilde{w}) h(\tilde{w})$, $\mathbb{E}(y^2 x_s^{2r} | \tilde{w}) h(\tilde{w})$ and $\mathbb{E}(y x_s^{3r} | \tilde{w}) h(\tilde{w})$ are twice continuously differentiable in \tilde{w} . Using these facts, and doing the usual change of variables, we can show that for large enough n

$$\begin{aligned} var\{\check{A}_n(\tilde{x}_0)\} &\doteq \frac{\mathfrak{R}_\kappa^{s-1}}{n a_n^{s-1}} x_{s,0}^{2r} \mathbb{E}(y^2 x_s^{2r} | \tilde{w}_0) h(\tilde{w}_0) \\ var\{\check{B}_n(\tilde{x}_0)\} &\doteq \frac{\mathfrak{R}_\kappa^{s-1}}{n a_n^{s-1}} \mathbb{E}(x_s^{4r} | \tilde{w}_0) h(\tilde{w}_0) \\ cov\{\check{A}_n(\tilde{x}_0), \check{B}_n(\tilde{x}_0)\} &\doteq \frac{\mathfrak{R}_\kappa^{s-1}}{n a_n^{s-1}} x_{s,0}^r \mathbb{E}(y x_s^{3r} | \tilde{w}_0) h(\tilde{w}_0). \end{aligned}$$

³Assumptions 2.1(iii) and (iv) ensure that the remainder terms in this Taylor expansion are well behaved. We avoid introducing any explicit remainder terms in this analysis as they do not affect the outcome of the paper.

Therefore, substituting these results in Equation B.1 and using the fact that $f(\tilde{x}_0) = x_{s,0}^r f(\tilde{w}_0, 1)$, a little algebra shows that

$$\begin{aligned} \text{var}\{\check{f}_{\text{proj}}(\tilde{x}_0)\} &\doteq \frac{\mathfrak{R}_\kappa^{s-1}}{na_n^{s-1}} \frac{x_{s,0}^{2r}}{\mathbb{E}^2(x_s^{2r}|\tilde{w}_0)h(\tilde{w}_0)} \{\mathbb{E}(y^2 x_s^{2r}|\tilde{w}_0) \\ &\quad + \mathbb{E}(x_{s,0}^{2r} f^2(\tilde{x})|\tilde{w}_0) - 2\mathbb{E}(y x_s^{2r} f(\tilde{x})|\tilde{w}_0)\} \\ &= \frac{\mathfrak{R}_\kappa^{s-1}}{na_n^{s-1}} \frac{x_{s,0}^{2r} \mathbb{E}(x_s^{2r} \{y - f(\tilde{x})\}^2|\tilde{w}_0)}{\mathbb{E}^2(x_s^{2r}|\tilde{w}_0)h(\tilde{w}_0)} \\ &= \frac{\mathfrak{R}_\kappa^{s-1}}{na_n^{s-1}} \frac{x_{s,0}^{2r} \mathbb{E}(x_s^{2r} \varepsilon^2|\tilde{w}_0)}{\mathbb{E}^2(x_s^{2r}|\tilde{w}_0)h(\tilde{w}_0)}. \end{aligned}$$

□

Lemma B.2. $\text{var}\{\check{f}_{\text{dir}}(\tilde{x}_0)\} \doteq \frac{\mathfrak{R}_\kappa^{s-1}}{na_n^{s-1}} \frac{x_{s,0}^{2r} \mathbb{E}(x_s^{-2r} \varepsilon^2|\tilde{w}_0)}{h(\tilde{w}_0)}$.

Proof. Similar to the proof of Lemma B.1. □

Lemma B.3. $\text{bias}\{\check{f}_{\text{proj}}(\tilde{x}_0)\} \doteq 0.5a_n^2 \mu_\kappa^2 x_{s,0}^r \text{tr} \left\{ \frac{\partial^2 f(\tilde{w}_0, 1)}{\partial \tilde{w} \partial \tilde{w}'} + \frac{2}{\mathbb{E}(x_s^{2r}|\tilde{w}_0)h(\tilde{w}_0)} \frac{\partial f(\tilde{w}_0, 1)}{\partial \tilde{w}} \frac{\partial \mathbb{E}(x_s^{2r}|\tilde{w}_0)h(\tilde{w}_0)}{\partial \tilde{w}'} \right\}$.

Proof. As in the proof of Lemma B.1, a Taylor expansion shows that

$$\begin{aligned} \mathbb{E}\check{f}_{\text{proj}}(\tilde{x}_0) - f(\tilde{x}_0) &\doteq \frac{1}{B(\tilde{x}_0)} \{\mathbb{E}\check{A}_n(\tilde{x}_0) - A(\tilde{x}_0)\} \\ &\quad - \frac{f(\tilde{x}_0)}{B(\tilde{x}_0)} \{\mathbb{E}\check{B}_n(\tilde{x}_0) - B(\tilde{x}_0)\}. \end{aligned}$$

But as (i) $\tilde{w}_0 \in \text{int}(S_{\tilde{w}})$ (ii) $\mathbb{E}(y x_s^r|\tilde{w})h(\tilde{w})$ and $\mathbb{E}(x_s^{2r}|\tilde{w})h(\tilde{w})$ are twice continuously differentiable in \tilde{w} , a change of variables approach yields that for large enough n

$$\begin{aligned} \mathbb{E}\check{A}_n(\tilde{x}_0) - A(\tilde{x}_0) &\doteq 0.5a_n^2 x_{s,0}^r \int_{S_\mathcal{K}} \tilde{u}' \frac{\partial^2 \{\mathbb{E}(y x_s^r|\tilde{w}_0)h(\tilde{w}_0)\}}{\partial \tilde{w} \partial \tilde{w}'} \tilde{u} \mathcal{K}(\tilde{u}) d\tilde{u} \\ \mathbb{E}\check{B}_n(\tilde{x}_0) - B(\tilde{x}_0) &\doteq 0.5a_n^2 \int_{S_\mathcal{K}} \tilde{u}' \frac{\partial^2 \{\mathbb{E}(x_s^{2r}|\tilde{w}_0)h(\tilde{w}_0)\}}{\partial \tilde{w} \partial \tilde{w}'} \tilde{u} \mathcal{K}(\tilde{u}) d\tilde{u}. \end{aligned}$$

Since it is easy to see that for $\eta \in C^2(S_{\tilde{w}})$

$$\int_{S_\mathcal{K}} \tilde{u}' \frac{\partial^2 \eta(\tilde{w})}{\partial \tilde{w} \partial \tilde{w}'} \tilde{u} \mathcal{K}(\tilde{u}) d\tilde{u} = \mu_\kappa^2 \text{tr} \left\{ \frac{\partial^2 \eta(\tilde{w})}{\partial \tilde{w} \partial \tilde{w}'} \right\},$$

the expressions for the bias of \check{A}_n and \check{B}_n reduce to

$$\begin{aligned} \mathbb{E}\check{A}_n(\tilde{x}_0) - A(\tilde{x}_0) &\doteq 0.5a_n^2 x_{s,0}^r \mu_\kappa^2 \text{tr} \left\{ \frac{\partial^2 \{\mathbb{E}(y x_s^r|\tilde{w}_0)h(\tilde{w}_0)\}}{\partial \tilde{w} \partial \tilde{w}'} \right\} \\ \mathbb{E}\check{B}_n(\tilde{x}_0) - B(\tilde{x}_0) &\doteq 0.5a_n^2 \mu_\kappa^2 \text{tr} \left\{ \frac{\partial^2 \{\mathbb{E}(x_s^{2r}|\tilde{w}_0)h(\tilde{w}_0)\}}{\partial \tilde{w} \partial \tilde{w}'} \right\}. \end{aligned}$$

Some algebra, and the fact that $f(\tilde{x}_0) = x_{s,0}^r f(\tilde{w}_0, 1)$, now leads to the desired result. □

Lemma B.4. $bias\{\check{f}_{\text{dir}}(\tilde{x}_0)\} \doteq 0.5a_n^2\mu_\kappa^2x_{s,0}^r \text{tr} \left\{ \frac{\partial^2 f(\tilde{w}_0,1)}{\partial \tilde{w} \partial \tilde{w}'} + \frac{2}{h(\tilde{w}_0)} \frac{\partial f(\tilde{w}_0,1)}{\partial \tilde{w}} \frac{\partial h(\tilde{w}_0)}{\partial \tilde{w}'} \right\}$.

Proof. Similar to the proof of Lemma B.3. \square

REFERENCES

- ARROW, K., H. CHENERY, B. MINHAS, AND R. SOLOW (1961): "Capital-Labor Substitution and Economic Efficiency," *The Review of Economics and Statistics*, 43, 225–250.
- BIERENS, H. J. (1985): "Kernel Estimators of regression functions," in *Nonparametric and semiparametric methods in econometrics and statistics. Proceedings of the fifth international symposium in Economic Theory and Econometrics*, ed. by W. A. Barnett, J. Powell, and G. Tauchen, pp. 99–144. Cambridge University Press.
- FAN, J. (1992): "Design Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.
- GALLANT, A. R. (1981): "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form," *Journal of Econometrics*, 15, 211–246.
- HÄRDLE, W. (1989): *Applied Nonparametric Regression*. Cambridge University Press.
- HÄRDLE, W., AND O. LINTON (1994): "Applied Nonparametric Methods," in *Handbook of Econometrics, vol. IV*, ed. by R. Engle, and D. McFadden. Elsevier Science B.V., pp. 2295–2339.
- HÄRDLE, W., AND E. MAMMEN (1993): "Comparing Nonparametric versus Parametric Regression Fits," *Annals of Statistics*, 21, 1926–1947.
- HILDRETH, C. (1954): "Point Estimates of Ordinates of Concave Functions," *Journal of the American Statistical Association*, 49, 598–619.
- MATZKIN, R. (1994): "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics, vol. IV*, ed. by R. Engle, and D. McFadden. Elsevier Science B.V., pp. 2524–2558.
- McFADDEN, D. L. (1984): "Econometric Analysis of Qualitative Response Models," in *Handbook of Econometrics, vol. II*, ed. by Z. Griliches, and M. Intriligator. Elsevier Science B.V., pp. 1396–1457.
- NEWBY, W. K. (1994): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233–253.
- PAGAN, A., AND A. ULLAH (1999): *Nonparametric Econometrics*. Cambridge University Press.
- RUPPERT, D., AND M. WAND (1994): "Multivariate Locally Weighted Least Squares Regression," *Annals of Statistics*, 22, 1346–1370.
- RUUD, P. A. (1997): "Restricted Least Squares Subject to Monotonicity and Concavity Constraints," in *Advances in Economics and Econometrics: Theory and Applications. Proceedings of the Seventh World Congress*, ed. by D. M. Kreps, and K. F. Wallis, vol. 3, pp. 166–187. Cambridge University Press.
- RYU, H. K. (1993): "Maximum Entropy Estimation of Density and Regression Functions," *Journal of Econometrics*, 56, 397–440.
- SCHUSTER, E. F. (1972): "Joint asymptotic distribution of the estimated regression function at a finite number of distinct points," *Annals of Mathematical Statistics*, 43, 84–88.
- TRIPATHI, G. (1999): "Local Semiparametric Efficiency Bounds Under Shape Restrictions," Accepted for publication in *Econometric Theory*.
- YATCHEW, A. (1988): "Some tests of nonparametric regression models," in *Dynamic Econometric Modelling. Proceedings of the third international symposium in Economic Theory and Econometrics*, ed. by W. Barnett, E. Berndt, and H. White, pp. 121–135. Cambridge University Press.
- (1998): "Nonparametric Regression Techniques in Economics," *Journal of Economic Literature*, 36, 669–721.

YATCHEW, A., AND L. BOS (1997): "Nonparametric Least Squares Regression and Testing in Economic Models," *Journal of Quantitative Economics*, 13(2), 81–31.

DEPARTMENT OF ECONOMICS, UNIVERSITY OF WISCONSIN, MADISON, WI-53706, USA.

E-mail address: `gtripath@ssc.wisc.edu`

INSTITUT FÜR STATISTIK UND ÖKONOMETRIE, HUMBOLDT-UNIVERSITÄT ZU BERLIN, D-10178 BERLIN, GERMANY.

E-mail address: `woocheol@wiwi.hu-berlin.de`