Econometric Theory, **19**, 2003, 640–663. Printed in the United States of America.
DOI: 10.1017/S026646660319408X

NONPARAMETRIC ESTIMATION OF HOMOGENEOUS FUNCTIONS

GAUTAM TRIPATHI
University of Wisconsin-Madison

WOOCHEOL KIM
Humboldt-Universität zu Berlin

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 f is homogeneous of known degree r. Using a local linear approach, we examine two ways of nonparametrically estimating f: (i) a "direct" 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. 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. Beginning with the pioneering paper of Hildreth (1954), much work has been done in this area. See, for example, Gallant (1981), Yatchew (1988), Härdle (1989, Ch. 8), Ryu (1993), Matzkin (1994), Ruud (1997), and Yatchew and Bos (1997). Readers unfamiliar with nonparametric estimation techniques relevant to econometrics are referred to 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 that happens to be homogeneous of known degree r. Recall that $f: S \to \mathbb{R}$ is said to be homogeneous of degree $r \in \mathbb{R}$ if $f(\lambda\tilde{x}) = \lambda^r f(\tilde{x})$ for all $(\lambda, \tilde{x}) \in \mathbb{R}_{++} \times S$ such that $\lambda\tilde{x} \in S$. Such functional forms are frequently encountered in microeconomic theory. For instance, the profit (resp. cost) function for a profit max-

We thank Don Andrews and an anonymous referee for comments that greatly improved this paper. The first author thanks Professor Wolfgang Härdle for hospitality 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 Sonderforschungsbereich 373 ("Quantifikation und Simulation Ökonomischer Prozesse") and the NSF via grants SES-0111917 and SES-0214081 is also gratefully acknowledged. Address correspondence to: Gautam Tripathi, Department of Economics, University of Wisconsin–Madison, Madison, WI 53706, USA; e-mail: gtripath@ssc.wisc.edu; and Woocheol Kim, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin, D-10178 Berlin, Germany; e-mail: woocheol@wiwi.hu-berlin.de.

  0266-4666/03 $12.00

imizing (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 that are homogeneous of degree one, that is, that exhibit constant returns to scale. See, for instance, the classic paper by Arrow, Chenery, Minhas, and Solow (1961). Labor economists often assume that the matching function, which relates the number of jobs formed during a certain period of time to the number of vacancies available during that period and some other variables, is linearly homogeneous. See, for example, Petrongolo and Pissarides (2001) and the references therein.

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 when 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 simplifies 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 though not an impossible task. In contrast, 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.

In the fully nonparametric case perhaps the simplest way of imposing homogeneity is to use a “direct” approach. In this approach we pick one variable as the numeraire and use it to normalize all variables. Estimation is then carried out using the normalized variables. For instance, Ruud (1997, p. 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 this approach when asked to nonparametrically estimate a homogeneous conditional expectation. Because we are so used to working with ratios of variables such as relative prices, 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 estimators obtained. Their analytical simplicity and ease of use should make the proposed estimators a useful addition to the tool kit 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 projection based approaches. In Section 4 we compare the asymptotic performance of the direct and projection based estimators and show how the error term conditional variance determines which estimator is better. Section 5 describes the results of a small simulation experiment, and in Section 6 we discuss some additional efficiency related issues. Section 7 concludes. All proofs are confined to the Appendixes.

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, $\tilde{x} = (x_1, \ldots, x_s)'$, $\widetilde{w} = (x_1/x_s, \ldots, x_{s-1}/x_s)'$, $\tilde{x}_j = (x_{1,j}, \ldots, x_{s,j})'$, and $\widetilde{w}_j = (x_{1,j}/x_{s,j}, \ldots, x_{s-1,j}/x_{s,j})'$. Here $S_{\tilde{x}}$ is a compact subset of $\mathbb{R}^s$ such that x_s (the last component of $\tilde{x} \in S_{\tilde{x}}$) is positive and bounded away from zero. The map $H_0 : S_{\tilde{x}} \to \mathbb{R}^{s-1}$ is the homogeneous of degree zero transformation $H_0(\tilde{x}) = \widetilde{w}$, and $S_{\widetilde{w}} = H_0(S_{\tilde{x}})$ is the image of $S_{\tilde{x}}$ under H_0. We use $\tilde{x}_0$ to denote a point that is fixed in $int(S_{\tilde{x}})$, and $\widetilde{w}_0 = H_0(\tilde{x}_0)$ denotes its image under H_0. Because the map $\tilde{x} \mapsto (\widetilde{w}, x_s)$ is one to one and continuous on $S_{\tilde{x}}$, it is straightforward to verify that $\widetilde{w}_0 \in int(S_{\widetilde{w}})$. The expression $C^k(int(S_{\tilde{x}}))$ is the set of all real valued functions on $int(S_{\tilde{x}})$ that have continuous partial derivatives up to order k. We say that $f \in C^k(S_{\tilde{x}})$ if $f \in C^k(int(S_{\tilde{x}}))$ and f, including all its partial derivatives up to order k, can be extended continuously to $S_{\tilde{x}}$. Finally, $L^2(S_{\tilde{x}})$ is the set of all square integrable functions on $S_{\tilde{x}}$ that are integrable with respect to the probability distribution on $S_{\tilde{x}}$, and $\mathcal{F}_r$ (resp. $\mathcal{G}_r$) is the set of all functions in $C^2(S_{\tilde{x}})$ (resp. $L^2(S_{\tilde{x}})$) that are also homogeneous of degree r. The symbol $\mathrel{\dot{=}}$ indicates "approximate equality," that is, equality modulo an additive but asymptotically negligible term. Unless stated otherwise, all limits are taken as the sample size $n \to \infty$.

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 independent and identically distributed (i.i.d.) random variables in $\mathbb{R} \times S_{\tilde{x}}$, and $\mathbb{E}(\varepsilon_j | \tilde{x}_j) = 0$.

(ii) The functional form of $f \in \mathcal{F}_r$ is unknown, but we do know r.

https://doi.org/10.1017/S026646660319408X Published online by Cambridge University Press

(iii) The conditional pdf of $(y, x_s|\widetilde{w})$ is twice continuously differentiable at $\widetilde{w}_0$.
(iv) $h(\widetilde{w})$, the pdf of $\widetilde{w} = H_0(\tilde{x})$, is twice continuously differentiable at $\widetilde{w}_0$ and $h(\widetilde{w}_0) > 0$.
(v) For some $\gamma > 0$, the map $\widetilde{w} \mapsto \mathbb{E}(|\varepsilon x_s^r|^{2+\gamma}|\widetilde{w})$ is bounded and continuous at $\widetilde{w}_0$.

The restrictions on $S_{\tilde{x}}$, namely, that $S_{\tilde{x}}$ is compact and that the last coordinate of $\tilde{x} \in S_{\tilde{x}}$ is positive and bounded away from zero, ensure that the conditional expectations in Lemmas 3.1 and 3.2 in Section 3 exist for all $r \in \mathbb{R}$. This allows us to handle any degree of homogeneity. The assumption that we know r is quite weak as economic theory frequently predicts the degree of homogeneity.[1] Because $S_{\tilde{x}}$ is compact, (ii) implies that f is also an element of $\mathcal{G}_r$. (iii) implies that $\mathbb{E}(yx_s^r|\widetilde{w}), \mathbb{E}(y^2x_s^{2r}|\widetilde{w})$, and $\mathbb{E}(yx_s^{3r}|\widetilde{w})$ are twice continuously differentiable at $\widetilde{w}_0$. This is used in the proof of Lemma B.1 in Appendix B. We use (iv) to ensure that the remainder terms in the Taylor expansions employed in Appendix A are well behaved. (iv) can be made more palatable if we interpret it to mean that we should carry out estimation and inference in regions where the density is bounded away from zero.[2] (v) provides sufficient moments so that we can prove the asymptotic normality of estimators of f.

3. ESTIMATION HEURISTICS

Because f is homogeneous of degree r, we can write

$$y = x_s^r f(\widetilde{w},1) + \varepsilon \Leftrightarrow \frac{y}{x_s^r} = f(\widetilde{w},1) + \frac{\varepsilon}{x_s^r}. \qquad \textbf{(1)}$$

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 estimator of $f(\tilde{x})$ based on the first representation as a "projection based" estimator, whereas the estimator based on the second formulation is called a "direct" estimator. 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 because 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 show that if $\text{var}(\varepsilon|\tilde{x})$ is homogeneous of degree zero (which includes homoskedasticity as a special case) then the projection based estimator is asymptotically better than the direct estimator, whereas if $\text{var}(\varepsilon|\tilde{x})$ is homogeneous of degree $2r \neq 0$ then the latter dominates the former.

Notice that if $r = 0$, that is, if we are estimating a homogeneous function of degree zero (e.g., a demand function), the two approaches will yield identical results. Furthermore, if $s = 1$ the problem is uninteresting because homogeneous functions are known up to scale in the one-dimensional case (because

https://doi.org/10.1017/S026646660319408X Published online by Cambridge University Press

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 $s > 1$.

Our estimation strategy is to approximate sample analogs of optimization problems that identify $f(\tilde{x})$ using the two representations in (1). At the population level we can use the first representation to write $f(\tilde{x})$ as $x_s^r \beta_p(\tilde{x})$, where β_p is identified as

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

Because $x_s^r \beta(\widetilde{w})$ is a homogeneous function of degree r for all $\beta \in \mathrm{L}^2(S_{\widetilde{w}})$, we can characterize $x_s^r \beta_p(\widetilde{w})$ as the orthogonal projection of y onto $\mathcal{G}_r$ using the usual L^2 inner product $\langle u,v \rangle_{\mathrm{L}^2} = \mathbb{E}\{uv\}$. In particular, we can use Lemma B.2 in Tripathi (2000) to show that this projection can be explicitly calculated as $x_s^r \mathbb{E}(y x_s^r | \widetilde{w}) / \mathbb{E}(x_s^{2r} | \widetilde{w})$. This explains the term *projection based* in describing an estimate obtained by using the first representation in (1). Similarly, a population level specification of $f(\tilde{x})$ using the second representation can be written as $x_s^r \beta_d(\widetilde{w})$, where we identify β_d as

$$\beta_d = \underset{\{\beta : \beta \in \mathrm{L}^2(S_{\widetilde{w}})\}}{\operatorname{argmin}} \mathbb{E}\left\{ \frac{y}{x_s^r} - \beta(\widetilde{w}) \right\}^2. \tag{3}$$

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

The preceding discussion shows that to estimate f, it suffices to estimate β_p and β_d. Because 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 $\tilde{x}_0$. In particular, let us now see how we can estimate $\beta_p(\widetilde{w}_0)$ and $\beta_d(\widetilde{w}_0)$.

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

$$\beta(\widetilde{w}) \mathrel{\mathring{=}} \beta(\widetilde{w}_0) + \nabla\beta(\widetilde{w}_0)'(\widetilde{w} - \widetilde{w}_0). \tag{4}$$

The unknown coefficients $\{\beta(\widetilde{w}_0), \nabla\beta(\widetilde{w}_0)\}$ can be estimated by doing least squares on sample analogs of (2) and (3) provided we can maintain the quality of the linear approximation in (4). This can be achieved by employing the usual device of local weighting to ensure that when estimating $\{\beta(\widetilde{w}_0), \nabla\beta(\widetilde{w}_0)\}$, the $\widetilde{w}$'s closer to $\widetilde{w}_0$ are given more weight than those observations that are farther away from $\widetilde{w}_0$. Following this approach, $\{\beta_p(\widetilde{w}_0), \nabla\beta_p(\widetilde{w}_0)\}$ and $\{\beta_d(\widetilde{w}_0), \nabla\beta_d(\widetilde{w}_0)\}$ can be estimated as

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

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

where $\tilde{b} = (b_0, b_1, \ldots, b_{s-1})$ and $\tilde{b}_* = (b_1, \ldots, b_{s-1})$.

Here $\hat{\beta}_p(\widetilde{w}_0)$ (resp. $\hat{\beta}_d(\widetilde{w}_0)$) is the projection based (resp. direct) local linear estimator of $f(\widetilde{w}_0,1)$. Similarly, $\widehat{\nabla\beta}_p(\widetilde{w}_0)$ (resp. $\widehat{\nabla\beta}_d(\widetilde{w}_0)$) is the projection based (resp. direct) local linear estimator of its gradient $\nabla f(\widetilde{w}_0,1)$. The kernel $\mathcal{K}$ (defined on $\mathbb{R}^{s-1}$) and the bandwidth a_n used previously satisfy the following conditions.

Assumption 3.1. $\mathcal{K}(\widetilde{w}) = \Pi_{i=1}^{s-1}\kappa(w_i)$, where $\kappa : \mathbb{R} \to \mathbb{R}$ is a continuous density with support $[-1,1]$ and is symmetric around zero. We also define $\Re_\kappa = \int_{-1}^{1} \kappa^2(u)\,du$, $\mu_{\kappa,2} = \int_{-1}^{1} u^2\kappa(u)\,du$, and $S_{\mathcal{K}} = [-1,1]^{s-1}$.

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

The asymptotic behavior of $\hat{\beta}$ and $\widehat{\nabla\beta}$ is given by the following results.

LEMMA 3.1. *Let Assumptions 2.1, 3.1, and 3.2 hold. Then*[3]

$$\begin{bmatrix} \sqrt{na_n^{s-1}}\{\hat{\beta}_p(\widetilde{w}_0) - f(\widetilde{w}_0,1) - bias_1\} \\ \sqrt{na_n^{s+1}}\{\widehat{\nabla\beta}_p(\widetilde{w}_0) - \nabla f(\widetilde{w}_0,1) - bias_2\} \end{bmatrix} \xrightarrow{d} \mathrm{N}(\widetilde{0}_{s\times 1}, \Sigma_p), \quad \textit{where}$$

$$\begin{bmatrix} bias_1 \\ bias_2 \end{bmatrix} = \begin{bmatrix} \dfrac{a_n^2}{2}\,\mu_{\kappa,2}\,\mathrm{tr}\{\nabla^2 f(\widetilde{w}_0,1)\} \\ \dfrac{a_n}{2\mu_{\kappa,2}} \displaystyle\int_{S_{\mathcal{K}}} \tilde{u}\mathcal{K}(\tilde{u})\tilde{u}'\,\nabla^2 f(\widetilde{w}_0,1)\,\tilde{u}\,d\tilde{u} \end{bmatrix}_{s\times 1} \quad \textit{and}$$

$$\Sigma_p = \begin{bmatrix} \dfrac{\mathbb{E}(x_s^{2r}\varepsilon^2|\widetilde{w}_0)\Re_\kappa^{s-1}}{\mathbb{E}^2(x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0)} & \widetilde{0}'_{(s-1)\times 1} \\ \widetilde{0}_{(s-1)\times 1} & \dfrac{\mathbb{E}(x_s^{2r}\varepsilon^2|\widetilde{w}_0)\displaystyle\int_{S_{\mathcal{K}}} \tilde{u}\tilde{u}'\mathcal{K}^2(\tilde{u})\,d\tilde{u}}{\mathbb{E}^2(x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0)\mu_{\kappa,2}^2} \end{bmatrix}_{s\times s}.$$

LEMMA 3.2. *Let Assumptions 2.1, 3.1, and 3.2 hold. Then*

$$\begin{bmatrix} \sqrt{na_n^{s-1}}\{\hat{\beta}_d(\widetilde{w}_0) - f(\widetilde{w}_0,1) - bias_1\} \\ \sqrt{na_n^{s+1}}\{\widehat{\nabla\beta}_d(\widetilde{w}_0) - \nabla f(\widetilde{w}_0,1) - bias_2\} \end{bmatrix} \xrightarrow{d} \mathrm{N}(\widetilde{0}_{s\times 1}, \Sigma_d),$$

https://doi.org/10.1017/S026646660319408X Published online by Cambridge University Press

where bias_1 *and* bias_2 *are defined in Lemma 3.1, and*

$$\Sigma_d = \begin{bmatrix} \dfrac{\mathbb{E}(x_s^{-2r}\varepsilon^2 \,|\, \widetilde{w}_0)\Re_\kappa^{s-1}}{h(\widetilde{w}_0)} & \widetilde{0}'_{(s-1)\times 1} \\ \widetilde{0}_{(s-1)\times 1} & \dfrac{\mathbb{E}(x_s^{-2r}\varepsilon^2 \,|\, \widetilde{w}_0)\displaystyle\int_{S_\mathcal{K}} \tilde{u}\tilde{u}'\mathcal{K}^2(\tilde{u})\,d\tilde{u}}{h(\widetilde{w}_0)\mu_{\kappa,2}^2} \end{bmatrix}_{s\times s}.$$

We use $\hat{f}_{\text{proj}}(\tilde{x}_0) = x_{s,0}^r \hat{\beta}_p(\widetilde{w}_0)$ and $\hat{f}_{\text{dir}}(\tilde{x}_0) = x_{s,0}^r \hat{\beta}_d(\widetilde{w}_0)$ to denote the projection based and direct local linear estimates of $f(\tilde{x}_0)$. Notice that $\hat{f}_{\text{proj}}$ and $\hat{f}_{\text{dir}}$ are homogeneous of degree r by construction; that is, we have obtained homogeneity constrained nonparametric estimators of f. Another nice feature of using the local linear approach is that both $f(\widetilde{w}_0,1)$ 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 because they can be expressed in a weighted least squares framework. See the proof of Lemma 3.1 in Appendix A for details.

Local linear estimators of conditional mean functions, without any homogeneity restrictions, have been extensively studied. See, for instance, Fan (1992), Ruppert and Wand (1994), Gozala and Linton (2000), and the references therein. If instead of a first-order approximation in (4) we had taken an mth-order Taylor expansion of $\beta(\widetilde{w})$ around $\widetilde{w}_0$, where $m > 1$, we would have obtained mth-order local polynomial estimators of β_p and β_d. In our case such higher order approximations are unnecessary because a linear approximation suffices to compare the asymptotic *mse* of $\hat{\beta}_p(\widetilde{w}_0)$ and $\hat{\beta}_d(\widetilde{w}_0)$. Of course, we could also have obtained locally constant (better known as Nadaraya–Watson) estimators of $\beta_p(\widetilde{w}_0)$ and $\beta_d(\widetilde{w}_0)$ by considering the "zeroth"-order approximation $\beta(\widetilde{w}) \mathrel{\mathring{=}} \beta(\widetilde{w}_0)$ in (4). In fact, it is easy to show that the projection based and direct Nadaraya–Watson estimators of $f(\tilde{x}_0)$ are given by

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

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

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 3.1 and 3.2, they do not have any terms involving the first derivatives of $f(\widetilde{w},1)$:

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

Hence when local linear estimators are used, 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 $f(\tilde{x}_0)$,

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

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

Although in Appendix B we also show that $\operatorname{var}\{\check{f}_{\text{proj}}(\tilde{x}_0)\} = \operatorname{var}\{\hat{f}_{\text{proj}}(\tilde{x}_0)\}$ and $\operatorname{var}\{\check{f}_{\text{dir}}(\tilde{x}_0)\} = \operatorname{var}\{\hat{f}_{\text{dir}}(\tilde{x}_0)\}$, 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 the projection based estimator with the direct estimator in terms of asymptotic *mse*.

Finally, let $\hat{f}(\tilde{x}_0)$ denote the usual unrestricted local linear estimator of $f(\tilde{x}_0)$ in the regression model $y = f(\tilde{x}) + \varepsilon$, that is,

$$\{\hat{f}(\tilde{x}_0), \widehat{\nabla f}(\tilde{x}_0)\} = \underset{(b_0,\tilde{b})\in\mathbb{R}\times\mathbb{R}^s}{\operatorname{argmin}} \sum_{j=1}^n \{y_j - b_0 - \tilde{b}'(\tilde{x}_j - \tilde{x}_0)\}^2 \mathcal{H}\left(\frac{\tilde{x}_j - \tilde{x}_0}{b_n}\right), \qquad \textbf{(5)}$$

where $\mathcal{H}$ is an appropriate kernel on $\mathbb{R}^s$ and the bandwidth b_n is a sequence of positive numbers such that $b_n \to 0$ and $nb_n^s \to \infty$. Note that when f is homogeneous of degree r, $\hat{f}$ will converge at a slower rate to f than $\hat{f}_{\text{proj}}$ or $\hat{f}_{\text{dir}}$. 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 $\hat{f}_{\text{proj}}$ and $\hat{f}_{\text{dir}}$ with $\hat{f}$.

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

Following Lemmas 3.1 and 3.2, 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

$$\operatorname{var}\{\hat{f}_{\text{proj}}(\tilde{x}_0)\} = \frac{x_{s,0}^{2r}\,\mathbb{E}(x_s^{2r}\varepsilon^2|\widetilde{w}_0)\Re_\kappa^{s-1}}{\mathbb{E}^2(x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0)},$$

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

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

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

where $\sigma^2(\tilde{x}) = \operatorname{var}(\varepsilon|\tilde{x})$ is the conditional variance function. It does not seem possible (at least to us) to compare the two variances if $\sigma^2(\tilde{x})$ is completely unknown. However, we can 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) $\sigma^2(\tilde{x}) = \psi^2(\widetilde{w})$ for some unknown ψ; that is, the error terms are conditionally heteroskedastic such that $\sigma(\tilde{x})$ is homogeneous of degree zero in the covariates, or
(ii) $\sigma^2(\tilde{x}) = x_s^{2r}\psi^2(\widetilde{w})$ for some unknown ψ; that is, the error terms are conditionally heteroskedastic such that $\sigma(\tilde{x})$ is homogeneous of degree $r \neq 0$ in the covariates.

Notice that Assumption 4.1(i) is automatically satisfied if the error term is homoskedastic. As an example of a model where homoskedasticity 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 output and factor prices. Assuming that the prices are measured without error but there is measurement error in the observed profit, we can write $y = f(\tilde{x}) + \varepsilon$. Because 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 homoskedastic.

Although homoskedasticity 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 one, consider the following model, which is motivated by the discussion in McFadden (1984, p. 1406).

Example 4.2.

Let $\tilde{x}$ be the vector of observed output and factor prices and $f^*(\tilde{x};u)$ the unobserved profit function of a competitive firm. The term u denotes a firm specific random parameter that 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 that are unobserved by the economist but are used by the firm when making production decisions. Because 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. In particular, it is homogeneous of degree one. Moreover, we can also verify that $\mathbb{E}(\varepsilon|\tilde{x}) = 0$ and that $\sigma(\tilde{x})$ is homogeneous of degree one.

Now it is easy to see that

$$\text{Assumption 4.1(i)} \Rightarrow \begin{cases} \mathbb{E}\{x_s^{-2r}\varepsilon^2|\widetilde{w}\} = \psi^2(\widetilde{w})\mathbb{E}\{x_s^{-2r}|\widetilde{w}\} \\ \mathbb{E}\{x_s^{2r}\varepsilon^2|\widetilde{w}\} = \psi^2(\widetilde{w})\mathbb{E}\{x_s^{2r}|\widetilde{w}\}, \end{cases}$$

$$\text{Assumption 4.1(ii)} \Rightarrow \begin{cases} \mathbb{E}\{x_s^{-2r}\varepsilon^2|\widetilde{w}\} = \psi^2(\widetilde{w}) \\ \mathbb{E}\{x_s^{2r}\varepsilon^2|\widetilde{w}\} = \psi^2(\widetilde{w})\mathbb{E}\{x_s^{4r}|\widetilde{w}\}. \end{cases}$$

Hence the expressions for the asymptotic variances simplify to

$$\text{Assumption 4.1(i)} \Rightarrow \begin{cases} \operatorname{var}\{\hat{f}_{\text{dir}}(\tilde{x}_0)\} = \dfrac{\psi^2(\widetilde{w}_0)\Re_\kappa^{s-1}x_{s,0}^{2r}\mathbb{E}(x_s^{-2r}|\widetilde{w}_0)}{h(\widetilde{w}_0)} \\ \operatorname{var}\{\hat{f}_{\text{proj}}(\tilde{x}_0)\} = \dfrac{\psi^2(\widetilde{w}_0)\Re_\kappa^{s-1}x_{s,0}^{2r}}{\mathbb{E}(x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0)}, \end{cases}$$

$$\text{Assumption 4.1(ii)} \Rightarrow \begin{cases} \operatorname{var}\{\hat{f}_{\text{dir}}(\tilde{x}_0)\} = \dfrac{\psi^2(\widetilde{w}_0)\Re_\kappa^{s-1}x_{s,0}^{2r}}{h(\widetilde{w}_0)} \\ \operatorname{var}\{\hat{f}_{\text{proj}}(\tilde{x}_0)\} = \dfrac{\psi^2(\widetilde{w}_0)\Re_\kappa^{s-1}x_{s,0}^{2r}\mathbb{E}(x_s^{4r}|\widetilde{w}_0)}{\mathbb{E}^2(x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0)}. \end{cases}$$

Because by the Cauchy–Schwarz inequality

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

we get that

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

under Assumption 4.1(i), and

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

under Assumption 4.1(ii). As the asymptotic bias for the two estimators is identical, we have $mse\{\hat{f}_{\text{proj}}(\tilde{x}_0)\} \le mse\{\hat{f}_{\text{dir}}(\tilde{x}_0)\}$ under Assumption 4.1(i) and $mse\{\hat{f}_{\text{dir}}(\tilde{x}_0)\} \le mse\{\hat{f}_{\text{proj}}(\tilde{x}_0)\}$ 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 estimator to use is not obvious but depends upon the nature of the heteroskedasticity 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 $s = 2$. A number n of observations on y were generated from $y = f(x_1, x_2) + \sigma(x_1, x_2)\varepsilon$, where $x_1, x_2 \overset{d}{=} \text{UIID}[1,2]$ and ε was chosen independently of (x_1, x_2). A Gaussian kernel was used to obtain $\hat{f}_{\text{proj}}$, $\hat{f}_{\text{dir}}$, and $\hat{f}$. For the first two estimators the bandwidth used was $cn^{-1/5}$, whereas 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:

$$f_1(x_1, x_2) = 10\sqrt{x_1 x_2} \quad \text{and} \quad \varepsilon_1 \overset{d}{=} \text{N}(0, 0.75), \qquad \textbf{(Model 1)}$$

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

TABLE 1. Average *mse* over grid for Model 1 (Cobb–Douglas)

		$\sigma(x_1, x_2) = 1$			$\sigma(x_1, x_2) = x_2$		
n	Average *mse*	$c = 0.5$	$c = 1$	$c = 2$	$c = 0.5$	$c = 1$	$c = 2$
50	$\hat{f}_{\text{proj}}$	0.046	0.039	0.045	0.101	0.083	0.087
	$\hat{f}_{\text{dir}}$	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	$\hat{f}_{\text{proj}}$	0.023	0.023	0.029	0.050	0.044	0.048
	$\hat{f}_{\text{dir}}$	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 grid 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	$\hat{f}_{\text{proj}}$	0.122	0.120	0.139	0.275	0.230	0.266
	$\hat{f}_{\text{dir}}$	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	$\hat{f}_{\text{proj}}$	0.066	0.067	0.097	0.146	0.132	0.148
	$\hat{f}_{\text{dir}}$	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

The terms f_1 and f_2 represent a Cobb–Douglas and constant elasticity of substitution (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 forms for $\sigma(x_1, x_2)$ 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 1,000 replications.

Recall that our asymptotic results are about the pointwise behavior of the *mse*. But because 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.

As seen in the tables, the average *mse* is ranked according to our asymptotic results. Moreover, except in one case[4] this ranking does not change when c is varied. When ε is homoskedastic, $\hat{f}_{\text{proj}}$ dominates, although there does not seem to be a dramatic difference between $\hat{f}_{\text{proj}}$ and $\hat{f}_{\text{dir}}$ in terms of average *mse*. When the conditional variance of ε is homogeneous of degree one, $\hat{f}_{\text{dir}}$ has the smallest average *mse* although once again the difference between $\hat{f}_{\text{proj}}$ and $\hat{f}_{\text{dir}}$ is not very substantial. However, in each case the homogeneity constrained estimators clearly outperform $\hat{f}$, the unrestricted local linear estimator 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 simulation results, namely, that they are limited in nature and may vary if the underlying model parameters are changed.

6. DISCUSSION

In this section we address two efficiency related issues. First, recall that we can rank $\hat{f}_{\text{proj}}(\tilde{x}_0)$ and $\hat{f}_{\text{dir}}(\tilde{x}_0)$ in terms of the asymptotic *mse* if we have some information about the homogeneity of $\tilde{x} \mapsto \sigma(\tilde{x})$. Hence an obvious question is:

in the absence of such information about $\sigma(\tilde{x})$, can we do better by (say) looking at a linear combination of the two? The second issue is more subtle and concerns the role of the numeraire x_s (the last component of $\tilde{x}$) in the definition of $\hat{f}_{\text{proj}}(\tilde{x}_0)$ and $\hat{f}_{\text{dir}}(\tilde{x}_0)$. Statistically, choosing a particular element of $\tilde{x}$ as the numeraire matters. For example, Lemmas 3.1 and 3.2 show how the asymptotic bias and variance of $\hat{f}_{\text{proj}}(\tilde{x}_0)$ and $\hat{f}_{\text{dir}}(\tilde{x}_0)$ depend upon x_s. Assuming that we have more than one candidate for the numeraire, each particular choice leads to a different estimator (projection based or direct) for $f(\tilde{x}_0)$. Hence another interesting question is: how can we optimally combine these estimators to obtain an estimator that is invariant to numeraire choice?

To answer these questions, it helps to reformulate the estimation problem described in Section 3 in more general terms. So let $\delta(\tilde{x})$ be a nonnegative function and consider the following weighted local least squares problem:

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

Letting $\hat{f}_\delta(\tilde{x}_0) = x_{s,0}^r \hat{\beta}_\delta(\widetilde{w}_0)$, it is apparent that the preceding optimization problem leads to a family of local linear estimators for $f(\tilde{x}_0)$ that are indexed by δ. In particular, for $\delta(\tilde{x}) = 1$ we obtain the projection based estimator $\hat{f}_{\text{proj}}(\tilde{x}_0)$ whereas $\delta(\tilde{x}) = x_s^{-2r}$ yields the direct estimator $\hat{f}_{\text{dir}}(\tilde{x}_0)$.

Following the arguments in Appendix A leading up to (A.7), it can be easily shown that $\hat{f}_\delta(\tilde{x}_0)$ is asymptotically linear,[5] that is,

$$\begin{aligned}&\sqrt{na_n^{s-1}}\{\hat{f}_\delta(\tilde{x}_0) - f(\tilde{x}_0) - 0.5a_n^2\mu_{\kappa,2}x_{s,0}^r \operatorname{tr}[\nabla^2 f(\widetilde{w}_0,1)]\} \qquad \textbf{(7)}\\ &\quad = \frac{x_{s,0}^r}{\mathbb{E}\{x_s^{2r}\delta(\tilde{x})|\widetilde{w}_0\}h(\widetilde{w}_0)} \frac{1}{\sqrt{na_n^{s-1}}} \sum_{j=1}^{n} \varepsilon_j x_{s,j}^r \mathcal{K}\left(\frac{\widetilde{w}_j - \widetilde{w}_0}{a_n}\right)\delta(\tilde{x}_j) + o_p(1).\end{aligned}$$

This representation is quite useful. For example, using the influence function of $\hat{f}_\delta(\tilde{x}_0)$ in (7), a straightforward application of the central limit theorem reveals that

$$\begin{aligned}&\sqrt{na_n^{s-1}}\{\hat{f}_\delta(\tilde{x}_0) - f(\tilde{x}_0) - 0.5a_n^2\mu_{\kappa,2}x_{s,0}^r \operatorname{tr}[\nabla^2 f(\widetilde{w}_0,1)]\}\\ &\quad \xrightarrow{d} \mathrm{N}\left(0, \frac{x_{s,0}^{2r}\mathbb{E}\{x_s^{2r}\sigma^2(\tilde{x})\delta^2(\tilde{x})|\widetilde{w}_0\}\Re_\kappa^{s-1}}{\mathbb{E}^2\{x_s^{2r}\delta(\tilde{x})|\widetilde{w}_0\}h(\widetilde{w}_0)}\right). \qquad \textbf{(8)}\end{aligned}$$

Expression (8) shows that the asymptotic variance of $\hat{f}_\delta(\tilde{x}_0)$ depends upon the weight δ but the asymptotic bias does not. To get some intuition behind the former result, note that δ and $\mathcal{K}$ act on different arguments in the weighted local least squares problem. In particular, δ operates on $\tilde{x}_j$ whereas $\mathcal{K}$ operates on $((\widetilde{w}_j - \widetilde{w}_0)/a_n)$. The latter fact ensures that all expectations in the asymptotic variance of $\hat{f}_\delta$ are conditional on $\widetilde{w}_0$. Hence as $\tilde{x}$ is not predictable by $\widetilde{w}$, the weighting function $\delta(\tilde{x})$ survives in the limit. Clearly, this will not happen if $\mathcal{K}$ and δ have the same arguments. In particular, solving

$$\min_{(b_0,\tilde{b})\in\mathbb{R}\times\mathbb{R}^s} \sum_{j=1}^{n} \{y_j - b_0 - \tilde{b}'(\tilde{x}_j - \tilde{x}_0)\}^2 \mathcal{H}\left(\frac{\tilde{x}_j - \tilde{x}_0}{b_n}\right)\delta(\tilde{x}_j),$$

which is the δ-weighted version of (5), will yield local linear estimators of $f(\tilde{x}_0)$ whose asymptotic bias and variance are independent of δ.[6]

Let us determine the optimal δ that minimizes the asymptotic variance of $\hat{f}_\delta(\tilde{x}_0)$. Observe that because

$$\mathbb{E}^2\{x_s^{2r}\delta(\tilde{x})|\widetilde{w}_0\} \le \mathbb{E}\{x_s^{2r}\sigma^2(\tilde{x})\delta^2(\tilde{x})|\widetilde{w}_0\}\mathbb{E}\left\{\frac{x_s^{2r}}{\sigma^2(\tilde{x})}\,|\widetilde{w}_0\right\}$$

by the Cauchy–Schwarz inequality, we have

$$\operatorname{var}\{\hat{f}_\delta(\tilde{x}_0)\} \ge \frac{x_{s,0}^{2r}\,\Re_\kappa^{s-1}}{\mathbb{E}\left\{\dfrac{x_s^{2r}}{\sigma^2(\tilde{x})}\,|\widetilde{w}_0\right\}h(\widetilde{w}_0)} \quad \text{for all } \delta. \tag{9}$$

Now if we choose $\delta(\tilde{x}) = c\sigma^{-2}(\tilde{x})$, where c can be any arbitrary positive constant, then (8) reveals that

$$\operatorname{var}\{\hat{f}_{c\sigma^{-2}(\tilde{x})}(\tilde{x}_0)\} = \frac{x_{s,0}^{2r}\,\Re_\kappa^{s-1}}{\mathbb{E}\left\{\dfrac{x_s^{2r}}{\sigma^2(\tilde{x})}\,|\widetilde{w}_0\right\}h(\widetilde{w}_0)}. \tag{10}$$

Thus by (9) and (10), it follows that the optimal weight function is given by $\delta(\tilde{x}) = c\sigma^{-2}(\tilde{x})$. Furthermore, on comparing (6) with (9), we can see that $\hat{f}_{\text{proj}}(\tilde{x}_0)$ is asymptotically efficient under Assumption 4.1(i) whereas $\hat{f}_{\text{dir}}(\tilde{x}_0)$ is asymptotically efficient under Assumption 4.1(ii). Therefore, $\hat{f}_{\text{proj}}(\tilde{x}_0)$ (resp. $\hat{f}_{\text{dir}}(\tilde{x}_0)$) automatically incorporates the optimal weighting scheme when $\sigma(\tilde{x})$ is homogeneous of degree $r = 0$ (resp. $r \ne 0$). Hence it suffices to restrict attention to the projection based estimator when $\sigma(\tilde{x})$ is homogeneous of degree zero. Similarly, there is no loss of generality in only looking at the direct estimator when $\sigma(\tilde{x})$ is homogeneous of degree $r \ne 0$. This suggests that if we have no prior information about the homogeneity of $\sigma(\tilde{x})$, the best way to proceed may be to first estimate the conditional variance function and then do the weighted local linear regression described earlier.

Another way of improving on $\hat{f}_{\text{proj}}(\tilde{x}_0)$ and $\hat{f}_{\text{dir}}(\tilde{x}_0)$ is to consider an optimal linear combination of the two. So let $\tilde{f} = (\hat{f}_{\text{proj}}(\tilde{x}_0), \hat{f}_{\text{dir}}(\tilde{x}_0))$ and $\widetilde{\alpha} = (\alpha_1, \alpha_2) \in \mathbb{R}^2$ such that $\alpha_1 + \alpha_2 = 1$. Then $\widetilde{\alpha}'\tilde{f}$ denotes a linear combination of $\hat{f}_{\text{proj}}(\tilde{x}_0)$ and $\hat{f}_{\text{dir}}(\tilde{x}_0)$ that is consistent for $f(\tilde{x}_0)$. It remains to determine the optimal $\widetilde{\alpha}$ for which the asymptotic *mse* of $\widetilde{\alpha}'\tilde{f}$ is minimized. To do so, we need the joint distribution of $\tilde{f}$. Using (7) and the Cramér–Wold device, it is easy to show that

$$\begin{bmatrix} \sqrt{na_n^{s-1}}\{\hat{f}_{\text{proj}}(\tilde{x}_0) - f(\tilde{x}_0) - 0.5a_n^2\mu_{\kappa,2}x_{s,0}^r \operatorname{tr}[\nabla^2 f(\widetilde{w}_0,1)]\} \\ \sqrt{na_n^{s-1}}\{\hat{f}_{\text{dir}}(\tilde{x}_0) - f(\tilde{x}_0) - 0.5a_n^2\mu_{\kappa,2}x_{s,0}^r \operatorname{tr}[\nabla^2 f(\widetilde{w}_0,1)]\} \end{bmatrix} \xrightarrow{d} \mathrm{N}(\widetilde{0}_{s\times 1}, V),$$

https://doi.org/10.1017/S026646660319408X Published online by Cambridge University Press

where

$$V = \begin{bmatrix} v_p & v_{pd} \\ v_{pd} & v_d \end{bmatrix} = \begin{bmatrix} \dfrac{x_{s,0}^{2r}\mathbb{E}\{x_s^{2r}\sigma^2(\tilde{x})|\widetilde{w}_0\}\Re_\kappa^{s-1}}{\mathbb{E}^2\{x_s^{2r}|\widetilde{w}_0\}h(\widetilde{w}_0)} & \dfrac{x_{s,0}^{2r}\mathbb{E}\{\sigma^2(\tilde{x})|\widetilde{w}_0\}\Re_\kappa^{s-1}}{\mathbb{E}\{x_s^{2r}|\widetilde{w}_0\}h(\widetilde{w}_0)} \\ \dfrac{x_{s,0}^{2r}\mathbb{E}\{\sigma^2(\tilde{x})|\widetilde{w}_0\}\Re_\kappa^{s-1}}{\mathbb{E}\{x_s^{2r}|\widetilde{w}_0\}h(\widetilde{w}_0)} & \dfrac{x_{s,0}^{2r}\mathbb{E}\{x_s^{-2r}\sigma^2(\tilde{x})|\widetilde{w}_0\}\Re_\kappa^{s-1}}{h(\widetilde{w}_0)} \end{bmatrix}.$$

Because the asymptotic bias of $\tilde{\alpha}'\tilde{f}$ does not depend upon $\tilde{\alpha}$, minimizing its asymptotic *mse* reduces to minimizing $\tilde{\alpha}'V\tilde{\alpha}$. It is straightforward to verify that for any symmetric positive definite matrix $P_{p\times p}$,

$$\frac{P^{-1}\tilde{1}}{\tilde{1}'P^{-1}\tilde{1}} = \operatorname*{argmin}_{\{\tilde{\gamma}\in\mathbb{R}^p:\tilde{\gamma}'\tilde{1}=1\}} \tilde{\gamma}'P\tilde{\gamma}, \tag{11}$$

where $\tilde{1}$ denotes a conformable vector of ones. Hence using (11),

$$\begin{bmatrix} \dfrac{v_d - v_{pd}}{v_p + v_d - 2v_{pd}} \\ \dfrac{v_p - v_{pd}}{v_p + v_d - 2v_{pd}} \end{bmatrix} = \operatorname*{argmin}_{\{\tilde{\alpha}\in\mathbb{R}^2:\tilde{\alpha}'\tilde{1}=1\}} \tilde{\alpha}'V\tilde{\alpha}.$$

Therefore, the optimal linear combination of $\hat{f}_{\text{proj}}(\tilde{x}_0)$ and $\hat{f}_{\text{dir}}(\tilde{x}_0)$ is

$$\hat{q} = \frac{v_d - v_{pd}}{v_p + v_d - 2v_{pd}}\hat{f}_{\text{proj}}(\tilde{x}_0) + \frac{v_p - v_{pd}}{v_p + v_d - 2v_{pd}}\hat{f}_{\text{dir}}(\tilde{x}_0). \tag{12}$$

In general, $\hat{q}$ (which is also homogeneous of degree r) puts nonzero mass on $\hat{f}_{\text{proj}}(\tilde{x}_0)$ and $\hat{f}_{\text{dir}}(\tilde{x}_0)$. However, if $\sigma(\tilde{x})$ is homogeneous of degree zero, then $v_{pd} = v_p$ and the entire mass is put on the projection estimator. Similarly, if $\sigma(\tilde{x})$ is homogeneous of degree $r \neq 0$, then $v_{pd} = v_d$ and all mass is put on the direct estimator. In practice, $\hat{q}$ can be implemented by using a preliminary estimator of V.[7]

Finally, we construct estimators that are invariant to numeraire choice. So let $\mathcal{I} = \{1 \le i \le s : x_i$ is positive and bounded away from zero$\}$, that is, $\mathcal{I}$ denotes the set of all valid numeraire indices. For $i \in \mathcal{I}$, define $\widetilde{w}_j(i) = (x_{1,j}/x_{i,j},\ldots, x_{i-1,j}/x_{i,j}, x_{i+1,j}/x_{i,j},\ldots,x_{s,j}/x_{i,j})$, $\tilde{z}_0(i) = (x_{1,0}/x_{i,0},\ldots,x_{i-1,0}/x_{i,0},1,x_{i+1,0}/x_{i,0},\ldots,x_{s,0}/x_{i,0})$, and $\hat{f}_\delta(\tilde{x}_0;i) = x_{i,0}^r\hat{\beta}_\delta(\widetilde{w}_0(i))$. As in (7), we can show that

$$\begin{aligned} &\sqrt{na_n^{s-1}}\{\hat{f}_\delta(\tilde{x}_0;i) - f(\tilde{x}_0) - 0.5a_n^2\mu_{\kappa,2}x_{i,0}^r\operatorname{tr}[\nabla^2 f(\tilde{z}_0(i))]\} \\ &\quad = \frac{x_{i,0}^r}{\mathbb{E}\{x_i^{2r}\delta(\tilde{x})|\widetilde{w}_0(i)\}h(\widetilde{w}_0(i))}\frac{1}{\sqrt{na_n^{s-1}}} \\ &\qquad \times \sum_{j=1}^n \varepsilon_j x_{i,j}^r \mathcal{K}\left(\frac{\widetilde{w}_j(i) - \widetilde{w}_0(i)}{a_n}\right)\delta(\tilde{x}_j) + o_p(1). \end{aligned} \tag{13}$$

https://doi.org/10.1017/S026646660319408X Published online by Cambridge University Press

Notice that the choice of numeraire (in this case the ith component of $\tilde{x}$) influences the asymptotic bias in addition to the asymptotic variance of $\hat{f}_\delta(\tilde{x}_0;i)$.

Now let $\{\delta_i(\tilde{x}) : i \in \mathcal{I}\}$ be a collection of weight functions. For convenience, assume that $\mathcal{I}$ has cardinality T and let $\tilde{g}$, $\tilde{g}_0$, and $\tilde{\mu}$ denote $T \times 1$ vectors whose ith components (for $i \in \mathcal{I}$) are given by $\hat{f}_{\delta_i}(\tilde{x}_0;i)$, $f(\tilde{x}_0)$, and $0.5\lambda^{1/2}\mu_{\kappa,2}x_{i,0}^r \operatorname{tr}[\nabla^2 f(\tilde{z}_0(i))]$, respectively. We also let W be a $T \times T$ diagonal matrix such that its ith diagonal element is given by $x_{i,0}^{2r}\mathbb{E}\{x_i^{2r}\sigma^2(\tilde{x})\delta_i^2(\tilde{x})|\widetilde{w}_0(i)\}\Re_\kappa^{s-1}/\mathbb{E}^2\{x_i^{2r}\delta_i(\tilde{x})|\widetilde{w}_0(i)\}h(\widetilde{w}_0(i))$, where $i \in \mathcal{I}$.

Using (13) and the Cramér–Wold device along with the fact that $na_n^{s+3} \to \lambda$, we can show that $\sqrt{na_n^{s-1}}(\tilde{g} - \tilde{g}_0) \xrightarrow{d} \mathrm{N}(\tilde{\mu}, W)$.[8] Next, let $\tilde{\alpha}$ be a $T \times 1$ vector such that its components sum to one, that is, $\sum_{i\in\mathcal{I}} \alpha_i = 1$. Then

$$\tilde{\alpha}'\tilde{g} = \sum_{i\in\mathcal{I}} \alpha_i \hat{f}_{\delta_i}(\tilde{x}_0;i)$$

denotes a consistent estimator of $f(\tilde{x}_0)$ that is invariant to numeraire choice. Therefore, the invariant estimator that minimizes asymptotic *mse* is given by $\hat{g} = \tilde{\alpha}_*'\tilde{g}$, where[9]

$$\tilde{\alpha}_* = \underset{\{\tilde{\alpha}\in R^T : \tilde{\alpha}'\tilde{1}=1\}}{\operatorname{argmin}} \tilde{\alpha}'(\tilde{\mu}\tilde{\mu}' + W)\tilde{\alpha} \overset{(11)}{=} \frac{(\tilde{\mu}\tilde{\mu}' + W)^{-1}\tilde{1}}{\tilde{1}'(\tilde{\mu}\tilde{\mu}' + W)^{-1}\tilde{1}}.$$

Here $\hat{g}$ yields the optimal projection based invariant estimator on setting $\delta_i(\tilde{x}) = 1$ for all $i \in \mathcal{I}$. Similarly, the optimal direct invariant estimator is obtained by setting $\delta_i(\tilde{x}) = x_i^{-2r}$ for $i \in \mathcal{I}$. Feasible versions can be implemented by using preliminary estimators of $\tilde{\mu}$ and W. One final point: the diagonal nature of W suggests that asymptotic variance can also be reduced by constructing an estimator based on more than one numeraire. However, because the asymptotic bias $\tilde{\mu}$ depends upon numeraire choice, such an estimator is not guaranteed to show any improvement in terms of asymptotic *mse*.

7. 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 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 homoskedastic.

(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.

If no prior information is available about the conditional variance function, the best way to proceed may be to first estimate the conditional variance function and then do a weighted local linear regression. In any case, a good empirical practice is to report estimates of f using the projection based, direct, and also the unrestricted local linear estimator of f. A large discrepancy between the reported results may indicate 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 by 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.

NOTES

1. Some well known examples are described earlier in Section 1.

2. Hengartner and Linton (1996) show that though nonparametric estimators of conditional mean functions remain asymptotically normal at points where the density of the conditioning variable is zero, their rate of convergence slows down and the constants associated with the limiting distribution change.

3. Here $\tilde{0}_{s\times 1}$ denotes a $s \times 1$ vector of zeros.

4. For Model 2 when $n = 100$ and $c = 2$. However, a decomposition of the average *mse* for this case revealed that although the average variance for $\hat{f}_{\text{dir}}$ was smaller than the average variance of $\hat{f}_{\text{proj}}$, the average squared bias for the former was bigger than the average squared bias for the latter. Therefore, although the ranking with respect to 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 $\hat{f}_{\text{dir}}$ caused the ranking to change.

5. To see this, replace $\mathcal{K}((\tilde{w}_j - \tilde{w}_0)/a_n)$ in the proof of Lemma 3.1 by $\mathcal{K}((\tilde{w}_j - \tilde{w}_0)/a_n)\delta(\tilde{x}_j)$.

6. Jones (1993) obtains a similar result for the asymptotic variance of a δ-weighted Nadaraya–Watson (i.e., locally constant) estimator. However, his results also show that the asymptotic bias of the weighted Nadaraya–Watson estimator does depend upon δ. As we have just shown, this does not hold for local linear estimators.

7. The optimal $\tilde{\alpha}$ in (12) depends upon the evaluation point $\tilde{x}_0$. This dependence can be eliminated by minimizing $\tilde{\alpha}'(\int_{\tilde{x}_0 \in S_{\tilde{x}}} V\, d\tilde{x}_0)\tilde{\alpha}$, the integrated version of $\tilde{\alpha}'V\tilde{\alpha}$. Hence the optimal linear combination that minimizes integrated asymptotic *mse* is given by

$$\frac{\int_{S_{\tilde{x}}}(v_d - v_{pd})\,d\tilde{x}_0}{\int_{S_{\tilde{x}}}(v_p + v_d - 2v_{pd})\,d\tilde{x}_0}\hat{f}_{\text{proj}}(\tilde{x}_0) + \frac{\int_{S_{\tilde{x}}}(v_p - v_{pd})\,d\tilde{x}_0}{\int_{S_{\tilde{x}}}(v_p + v_d - 2v_{pd})\,d\tilde{x}_0}\hat{f}_{\text{dir}}(\tilde{x}_0).$$

Numerical integration can be used to construct the feasible weights.

8. Because $\hat{\beta}_{\delta_i}(\tilde{w}_0(i))$ and $\hat{\beta}_{\delta_l}(\tilde{w}_0(l))$ are evaluated at different points when $i \neq l$, the asymptotic covariance between $\hat{f}_{\delta_i}(\tilde{x}_0; i)$ and $\hat{f}_{\delta_l}(\tilde{x}_0; l)$ is zero for $i \neq l$. Hence the asymptotic variance-covariance matrix W is diagonal.

9. Notice that $\tilde{\alpha}_*$ depends upon $\tilde{x}_0$. The optimal $\tilde{\alpha}$ that minimizes integrated asymptotic *mse* (and hence does not depend upon $\tilde{x}_0$) is given by $\{\int_{S_{\tilde{x}}}(\tilde{\mu}\tilde{\mu}' + W)d\tilde{x}_0\}^{-1}\tilde{1}/\tilde{1}'\{\int_{S_{\tilde{x}}}(\tilde{\mu}\tilde{\mu}' + W)d\tilde{x}_0\}^{-1}\tilde{1}$.

REFERENCES

Arrow, K., H. Chenery, B. Minhas, & R. Solow (1961) Capital-labor substitution and economic efficiency. *Review of Economics and Statistics* 43, 225–250.

Bierens, H.J. (1985) Kernel estimators of regression functions. In W.A. Barnett, J. Powell, & G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, pp. 99–144. Cambridge, UK: 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.

Gozalo, P. & O. Linton (2000) Local nonlinear least squares: Using parametric information in nonparametric regression. *Journal of Econometrics* 99, 63–106.

Härdle, W. (1989) *Applied Nonparametric Regression*. New York: Cambridge University Press.

Härdle, W. & O. Linton (1994) Applied nonparametric methods. In R. Engle, & D. McFadden (eds.), *Handbook of Econometrics*, vol. IV, pp. 2295–2339. Amsterdam: Elsevier Science B.V.

Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.

Hengartner, N.W. & O.B. Linton (1996) Nonparametric regression estimation at design poles and zeros. *Canadian Journal of Statistics* 24, 583–591.

Hildreth, C. (1954) Point estimates of ordinates of concave functions. *Journal of the American Statistical Association* 49, 598–619.

Jones, M. (1993) Do not weight for heteroscedasticity in nonparametric regression. *Australian Journal of Statistics* 35, 89–92.

Matzkin, R. (1994) Restrictions of economic theory in nonparametric methods. In R. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. IV, pp. 2524–2558. Elsevier Science B.V.

McFadden, D.L. (1984) Econometric analysis of qualitative response models. In Z. Griliches & M. Intriligator (eds.), *Handbook of Econometrics*, vol. II, pp. 1396–1457. Amsterdam: Elsevier Science B.V.

Pagan, A. & A. Ullah (1999) *Nonparametric Econometrics*. New York: Cambridge University Press.

Petrongolo, B. & C.A. Pissarides (2001) Looking into the black box: A survey of the matching function. *Journal of Economic Literature* 39, 390–431.

Ruppert, D. & 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 D.M. Kreps & K.F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications, Proceedings of the Seventh World Congress*, vol. 3, pp. 166–187. Cambridge, UK: 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. (2000) Local semiparametric efficiency bounds under shape restrictions. *Econometric Theory* 16, 729–739.

Yatchew, A. (1988) Some tests of nonparametric regression models. In W. Barnett, E. Berndt, & H. White (eds.), *Dynamic Econometric Modelling: Proceedings of the Third International Sympo-*

sium in Economic Theory and Econometrics, pp. 121–135. Cambridge, UK: Cambridge University Press.

Yatchew, A. (1998) Nonparametric regression techniques in economics. *Journal of Economic Literature* 36, 669–721.

Yatchew, A. & L. Bos (1997) Nonparametric least squares regression and testing in economic models. *Journal of Quantitative Economics* 13, 81–31.

APPENDIX A: TECHNICAL DETAILS—LOCAL LINEAR ESTIMATORS

Proof of Lemma 3.1. (The referee has pointed out that Lemmas 3.1 and 3.2 are special cases of Theorem 2 in Gozalo and Linton (2000). However, as some readers may find it instructive to see a direct proof, we provide one.) Throughout this proof let $Q = \text{diag}_{s\times s}[1, a_n, \dots, a_n]$ and $\Omega = \text{diag}_{n\times n}[\mathcal{K}((\widetilde{w}_1 - \widetilde{w}_0)/a_n), \dots, \mathcal{K}((\widetilde{w}_n - \widetilde{w}_0)/a_n)]$. Furthermore, we also define

$$\hat{\theta}(\widetilde{w}_0) = \begin{bmatrix} \hat{\theta}_0(\widetilde{w}_0) \\ \hat{\theta}_1(\widetilde{w}_0) \\ \vdots \\ \hat{\theta}_{s-1}(\widetilde{w}_0) \end{bmatrix} = \begin{bmatrix} \hat{\beta}_p(\widetilde{w}_0) \\ \widehat{\nabla\beta_p}(\widetilde{w}_0) \end{bmatrix}_{s\times 1}, \quad Z = \begin{bmatrix} x^r_{s,1} & x^r_{s,1}\left(\dfrac{\widetilde{w}'_1 - \widetilde{w}'_0}{a_n}\right) \\ \vdots & \vdots \\ x^r_{s,n} & x^r_{s,n}\left(\dfrac{\widetilde{w}'_n - \widetilde{w}'_0}{a_n}\right) \end{bmatrix}_{n\times s},$$

and $\tilde{y} = (y_1, \dots, y_n)$. Using this notation, it is straightforward to see that $\hat{\theta}(\widetilde{w}_0)$ is the solution to the following weighted least squares problem:

$$\min_{\tilde{b}\in\mathbb{R}^s} (\tilde{y} - ZQ\tilde{b})'\Omega(\tilde{y} - ZQ\tilde{b}).$$

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

$$\hat{\theta}(\widetilde{w}_0) = Q^{-1}(Z'\Omega Z)^{-1}Z'\Omega\tilde{y}.$$

Let us write $\hat{\theta}(\widetilde{w}_0) = Q^{-1}S^{-1}\tilde{t}$, where $S = (1/na_n^{s-1})Z'\Omega Z$ and $\tilde{t} = (1/na_n^{s-1})Z'\Omega\tilde{y}$. Straightforward calculations show that we can write S as the partitioned matrix $S = \begin{bmatrix} s_{00} & s'_{01} \\ s_{01} & s_{11} \end{bmatrix}$, where $s_{00} = (1/na_n^{s-1})\sum_{j=1}^n x^{2r}_{s,j}\mathcal{K}(\widetilde{w}_j - \widetilde{w}_0/a_n)$,

$$s_{01} = \frac{1}{na_n^{s-1}}\sum_{j=1}^n x^{2r}_{s,j}\left(\frac{\widetilde{w}_j - \widetilde{w}_0}{a_n}\right)\mathcal{K}\left(\frac{\widetilde{w}_j - \widetilde{w}_0}{a_n}\right), \quad \text{and}$$

$$s_{11} = \frac{1}{na_n^{s-1}}\sum_{j=1}^n x^{2r}_{s,j}\left(\frac{\widetilde{w}_j - \widetilde{w}_0}{a_n}\right)\left(\frac{\widetilde{w}'_j - \widetilde{w}'_0}{a_n}\right)\mathcal{K}\left(\frac{\widetilde{w}_j - \widetilde{w}_0}{a_n}\right).$$

Similarly, we write $\tilde{t} = \begin{bmatrix} t_0 \\ t_1 \end{bmatrix}$, where $t_0 = (1/na_n^{s-1})\sum_{j=1}^n y_j x^r_{s,j}\mathcal{K}((\widetilde{w}_j - \widetilde{w}_0)/a_n)$ and $t_1 = 1/na_n^{s-1}\sum_{j=1}^n y_j x^r_{s,j}((\widetilde{w}_j - \widetilde{w}_0)/a_n)\mathcal{K}((\widetilde{w}_j - \widetilde{w}_0)/a_n)$. But because $y_j = x^r_{s,j} f(\widetilde{w}_j, 1) + \varepsilon_j$, we can express $\tilde{t} = \tilde{\tau} + \tilde{t}^*$, where

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

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

Hence letting $\theta(\widetilde{w}_0) = \begin{bmatrix}f(\widetilde{w}_0,1)\\ \nabla f(\widetilde{w}_0,1)\end{bmatrix}$, we have that

$$\hat{\theta}(\widetilde{w}_0) - \theta(\widetilde{w}_0) = Q^{-1}S^{-1}\tilde{\tau} + Q^{-1}S^{-1}\tilde{t}^* - \theta(\widetilde{w}_0). \tag{A.2}$$

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

$$f(\widetilde{w}_j,1) = f(\widetilde{w}_0,1) + a_n\left(\frac{\widetilde{w}_j-\widetilde{w}_0}{a_n}\right)'\nabla f(\widetilde{w}_0,1)$$

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

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

$$\tilde{t}^* = SQ\theta(\widetilde{w}_0) + \frac{a_n^2}{2}\tilde{c} + \begin{bmatrix}o_p(a_n^2)\\ \vdots\\ o_p(a_n^2)\end{bmatrix}_{s\times 1}, \quad \text{where}$$

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

Because it is easy to verify that

$$S \xrightarrow{p} \begin{bmatrix}\mathbb{E}(x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0) & \tilde{0}'_{(s-1)\times 1}\\ \tilde{0}_{(s-1)\times 1} & \mu_{\kappa,2}\mathbb{E}(x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0)I_{s-1}\end{bmatrix}, \tag{A.3}$$

https://doi.org/10.1017/S026646660319408X Published online by Cambridge University Press

where I_{s-1} denotes the $(s-1)\times(s-1)$ identity matrix, we get that

$$Q^{-1}S^{-1}\tilde{t}^{*} - \theta(\widetilde{w}_0) = \frac{a_n^2}{2}\,Q^{-1}S^{-1}\tilde{c} + \begin{bmatrix} o_p(a_n^2) \\ o_p(a_n) \\ \vdots \\ o_p(a_n) \end{bmatrix}_{s\times 1}.$$

Let us further simplify the right-hand side of this equation. To do so, first use a law of large numbers to show that

$$\tilde{c} \xrightarrow{p} \mathbb{E}(x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0)\begin{bmatrix} \mathrm{vec}'(\mu_{\kappa,2}I_{s-1}) \\ \int_{S_{\mathcal{K}}} \tilde{u}\mathcal{K}(\tilde{u})\mathrm{vec}'(\tilde{u}\tilde{u}')\,d\tilde{u} \end{bmatrix}\mathrm{vec}(\nabla^2 f(\widetilde{w}_0,1)) \tag{A.4}$$

Next, using (A.3), (A.4), and a little matrix manipulation,

$$\frac{a_n^2}{2}\,Q^{-1}S^{-1}\tilde{c} = \begin{bmatrix} \frac{a_n^2}{2}\,\mu_{\kappa,2}\,\mathrm{tr}\{\nabla^2 f(\widetilde{w}_0,1)\} \\ \frac{a_n}{2\mu_{\kappa,2}}\int_{S_{\mathcal{K}}} \tilde{u}\mathcal{K}(\tilde{u})\tilde{u}'\,\nabla^2 f(\widetilde{w}_0,1)\,\tilde{u}\,d\tilde{u} \end{bmatrix} + \begin{bmatrix} o_p(a_n^2) \\ o_p(a_n) \\ \vdots \\ o_p(a_n) \end{bmatrix}_{s\times 1}.$$

Thus letting $\widetilde{bias} = \begin{bmatrix} (a_n^2/2)\,\mu_{\kappa,2}\,\mathrm{tr}\{\nabla^2 f(\widetilde{w}_0,1)\} \\ (a_n/2\mu_{\kappa,2})\int_{S_{\mathcal{K}}} \tilde{u}\mathcal{K}(\tilde{u})\tilde{u}'\,\nabla^2 f(\widetilde{w}_0,1)\,\tilde{u}\,d\tilde{u} \end{bmatrix}$, (A.2) reduces to

$$\hat{\theta}(\widetilde{w}_0) - \theta(\widetilde{w}_0) - \widetilde{bias} = Q^{-1}S^{-1}\tilde{\tau} + \begin{bmatrix} o_p(a_n^2) \\ o_p(a_n) \\ \vdots \\ o_p(a_n) \end{bmatrix}_{s\times 1}. \tag{A.5}$$

Now we show that $S^{-1}\tilde{\tau}$ is asymptotically normal. A straightforward application of the Lindeberg–Lyapunov central limit theorem for triangular arrays and the Cramér–Wold device reveals that under Assumption 2.1(v)

$$\sqrt{na_n^{s-1}}\,\tilde{\tau} \xrightarrow{d} \mathrm{N}(\tilde{0}_{s\times 1}, K\mathbb{E}(\varepsilon^2 x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0)), \tag{A.6}$$

where

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

Thus by (A.1), (A.3), (A.6), and Slutsky,

$$\sqrt{na_n^{s-1}}S^{-1}\tilde{\tau} = \frac{1}{\mathbb{E}(x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0)\sqrt{na_n^{s-1}}}\begin{bmatrix}\tau_0\\ \frac{\tau_1}{\mu_{\kappa,2}}\end{bmatrix} + \begin{bmatrix}o_p(1)\\ \vdots \\ o_p(1)\end{bmatrix}_{s\times 1} \xrightarrow{d} \mathrm{N}(\widetilde{0}_{s\times 1}, \Sigma_p).$$

Finally, premultiplying both sides of (A.5) by $\sqrt{na_n^{s-1}}Q$ and using the fact that the sequence na_n^{s+3} is bounded, we have

$$\sqrt{na_n^{s-1}}Q\{\hat{\theta}(\widetilde{w}_0) - \theta(\widetilde{w}_0) - \widetilde{bias}\} = \frac{1}{\mathbb{E}(x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0)\sqrt{na_n^{s-1}}}\begin{bmatrix}\tau_0\\ \frac{\tau_1}{\mu_{\kappa,2}}\end{bmatrix} + \begin{bmatrix}o_p(1)\\ \vdots \\ o_p(1)\end{bmatrix}_{s\times 1} \quad \textbf{(A.7)}$$

Therefore, it follows that

$$\sqrt{na_n^{s-1}}Q\{\hat{\theta}(\widetilde{w}_0) - \theta(\widetilde{w}_0) - \widetilde{bias}\} \xrightarrow{d} \mathrm{N}(\widetilde{0}_{s\times 1}, \Sigma_p).$$

But this is the desired result. ■

Proof of Lemma 3.2 In the proof of Lemma 3.1, replace $x_{s,j}^r$ by 1, y_j by $y_j/x_{s,j}^r$, and ε_j by $\varepsilon_j/x_{s,j}^r$. ■

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. $\mathrm{var}\{\check{f}_{\mathrm{proj}}(\tilde{x}_0)\} \doteq (\Re_\kappa^{s-1}/na_n^{s-1})(x_{s,0}^{2r}\mathbb{E}(x_s^{2r}\varepsilon^2|\widetilde{w}_0)/\mathbb{E}^2(x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0))$.

Proof of Lemma B.1. Observe that we can write $\check{f}_{\mathrm{proj}}(\tilde{x}_0) = \check{A}(\tilde{x}_0)/\check{B}(\tilde{x}_0)$ and $f(\tilde{x}_0) = A(\tilde{x}_0)/B(\tilde{x}_0)$, where

$$\check{A}(\tilde{x}_0) = \frac{x_{s,0}^r}{na_n^{s-1}}\sum_{j=1}^n y_j x_{s,j}^r \mathcal{K}\left(\frac{\widetilde{w}_0 - \widetilde{w}_j}{a_n}\right), \qquad A(\tilde{x}_0) = x_{s,0}^r\mathbb{E}(yx_s^r|\widetilde{w}_0)h(\widetilde{w}_0),$$

$$\check{B}(\tilde{x}_0) = \frac{1}{na_n^{s-1}}\sum_{j=1}^n x_{s,j}^{2r}\mathcal{K}\left(\frac{\widetilde{w}_0 - \widetilde{w}_j}{a_n}\right), \qquad \text{and} \quad B(\tilde{x}_0) = \mathbb{E}(x_s^{2r}|\widetilde{w})h(\widetilde{w}_0).$$

https://doi.org/10.1017/S026646660319408X Published online by Cambridge University Press

Then by a Taylor expansion (Assumptions 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),

$$\check{f}_{\text{proj}}(\tilde{x}_0) - \mathbb{E}\,\check{f}_{\text{proj}}(\tilde{x}_0) \doteq \frac{1}{B(\tilde{x}_0)}\{\check{A}(\tilde{x}_0) - \mathbb{E}\check{A}(\tilde{x}_0)\} - \frac{f(\tilde{x}_0)}{B(\tilde{x}_0)}\{\check{B}(\tilde{x}_0) - \mathbb{E}\check{B}(\tilde{x}_0)\}.$$

Therefore,

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

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

$$\operatorname{var}\{\check{A}(\tilde{x}_0)\} \doteq \frac{\Re_\kappa^{s-1}}{na_n^{s-1}}\, x_{s,0}^{2r}\,\mathbb{E}(y^2x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0),$$

$$\operatorname{var}\{\check{B}(\tilde{x}_0)\} \doteq \frac{\Re_\kappa^{s-1}}{na_n^{s-1}}\,\mathbb{E}(x_s^{4r}|\widetilde{w}_0)h(\widetilde{w}_0), \quad \text{and}$$

$$\operatorname{cov}\{\check{A}(\tilde{x}_0), \check{B}(\tilde{x}_0)\} \doteq \frac{\Re_\kappa^{s-1}}{na_n^{s-1}}\, x_{s,0}^{r}\,\mathbb{E}(yx_s^{3r}|\widetilde{w}_0)h(\widetilde{w}_0).$$

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

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

■

LEMMA B.2 $\operatorname{var}\{\check{f}_{\text{dir}}(\tilde{x}_0)\} \doteq (\Re_\kappa^{s-1}/na_n^{s-1})[x_{s,0}^{2r}\,\mathbb{E}(x_s^{-2r}\varepsilon^2|\widetilde{w}_0)/h(\widetilde{w}_0)]$.

Proof of Lemma B.2. Similar to the proof of Lemma B.1.

LEMMA B.3.

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

Proof of Lemma B.3. By a Taylor expansion,

$$\mathbb{E}\check{f}_{\text{proj}}(\tilde{x}_0) - f(\tilde{x}_0) \doteq \frac{1}{B(\tilde{x}_0)}\{\mathbb{E}\check{A}(\tilde{x}_0) - A(\tilde{x}_0)\} - \frac{f(\tilde{x}_0)}{B(\tilde{x}_0)}\{\mathbb{E}\check{B}(\tilde{x}_0) - B(\tilde{x}_0)\}.$$

But as $\widetilde{w}_0 \in int(S_{\widetilde{w}})$ and $\mathbb{E}(yx_s^r|\widetilde{w})h(\widetilde{w})$, $\mathbb{E}(x_s^{2r}|\widetilde{w})h(\widetilde{w})$ are twice continuously differentiable at $\widetilde{w}_0$, a change of variables yields that

$$\mathbb{E}\check{A}(\tilde{x}_0) - A(\tilde{x}_0) \doteq 0.5a_n^2x_{s,0}^r\int_{S_{\mathcal{K}}}\tilde{u}'\frac{\partial^2\{\mathbb{E}(yx_s^r|\widetilde{w}_0)h(\widetilde{w}_0)\}}{\partial\widetilde{w}\partial\widetilde{w}'}\tilde{u}\,\mathcal{K}(\tilde{u})\,d\tilde{u} \quad \text{and}$$

$$\mathbb{E}\check{B}(\tilde{x}_0) - B(\tilde{x}_0) \doteq 0.5a_n^2\int_{S_{\mathcal{K}}}\tilde{u}'\frac{\partial^2\{\mathbb{E}(x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0)\}}{\partial\widetilde{w}\partial\widetilde{w}'}\tilde{u}\,\mathcal{K}(\tilde{u})\,d\tilde{u}$$

for large enough n. Because it is easy to see that for any η

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

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

$$\mathbb{E}\check{A}(\tilde{x}_0) - A(\tilde{x}_0) \doteq 0.5a_n^2x_{s,0}^r\mu_{\kappa,2}\operatorname{tr}\left\{\frac{\partial^2[\mathbb{E}(yx_s^r|\widetilde{w}_0)h(\widetilde{w}_0)]}{\partial\widetilde{w}\partial\widetilde{w}'}\right\} \quad \text{and}$$

$$\mathbb{E}\check{B}(\tilde{x}_0) - B(\tilde{x}_0) \doteq 0.5a_n^2\mu_{\kappa,2}\operatorname{tr}\left\{\frac{\partial^2[\mathbb{E}(x_s^{2r}|\widetilde{w}_0)h(\widetilde{w}_0)]}{\partial\widetilde{w}\partial\widetilde{w}'}\right\}.$$

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

LEMMA B.4.

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

Proof of Lemma B.4. Similar to the proof of Lemma B.3.

https://doi.org/10.1017/S026646660319408X Published online by Cambridge University Press