Nonparametric Policy Analysis

JAMES H. STOCK*

This article considers the problem of predicting the mean effect of a change in the distribution of certain policy-related variables on a dependent variable (Y). This is conventionally done using a parametric model. If, however, the conditional expectation of Y, given policy and nonpolicy variables X, is unaltered by the policy intervention, and if the support of X after the policy intervention lies within the support of X before the intervention, then this analysis can be performed nonparametrically. The proposed nonparametric estimator is developed for the model $Y_j = g(X_j) + A'd_j + u_j$, where $g(\cdot)$ is a continuous unknown function of the continuous variables X, d_j is an m-vector of dummy variables, A is an m-vector of unknown parameters representing fixed cell-specific effects, and u_j is an error term with $E(u_j \mid x_j, d_j) = 0$. The estimand is $B = Eg(X_j^*) - Eg(X_j)$, where X and X^* (respectively) denote the values of X before and after the policy intervention. A nonparametric estimator B_n is proposed. The estimator is the sample average of the difference between the kernel regression estimates of $E(Y \mid X_j^*, d_j)$ and $E(Y \mid X_j, d_j)$. To estimate these conditional expectations, A is first estimated using the residuals from nonparametric regressions of Y and d on X. The consistency and asymptotic normality of B_n are studied. The estimator, along with two estimators of its variance, is examined in a Monte Carlo experiment. In this experiment, the cost of using the nonparametric estimator, relative to the efficient parametric estimator, is found to be modest in terms of increased root mean squared error. When the dimension of X is large and the sample size is small, however, the nonparametric estimator can exhibit substantial bias.

KEY WORDS: Benefits estimation; Kernel regression; Policy effects.

1. INTRODUCTION

A common econometric problem is predicting the average effect of a proposed policy on some dependent variable, where the variable typically is either directly or indirectly related to individual welfare. For example, an analyst might be interested in estimating the mean change in house prices resulting from cleaning up a local hazardous-waste site, the change in lost work days resulting from a reduction in air pollution, or the change in average consumption resulting from a change in income taxes. More generally, we can think of a policy as transforming some or all of the elements of a k-dimensional vector of independent variables from an original value, X, to X^*. A typical policy analysis problem is estimating the mean benefits of this shift, that is, the mean effect of this change in X on the dependent variable of interest, Y.

The usual econometric approach to this problem is to specify a parametric model relating the independent and dependent variables—for example, a linear regression function—and to estimate the parameters of this model using, say, least squares or maximum likelihood. The effects, or benefits, of the proposed policy are then predicted using the estimated parametric model. Unfortunately, although economic theory might suggest the appropriate set of independent variables, it often provides little guidance concerning the precise parametric model to estimate. The consequences of this ambiguity can be severe: If the parametric model is misspecified, the corresponding benefits estimator is in general inconsistent.

This article proposes a procedure for estimating the mean effect of certain types of policy interventions when theory gives little guidance about the functional form relating the independent and dependent variables. The estimator is developed for a semiparametric model that applies when the data are drawn from discrete observational cells. Specifically, it is assumed that the regression function can be expressed as an unknown function $g(x)$, where x is composed of continuous policy and control variables, plus cell-specific effects. This model reflects a compromise between a fully nonparametric model, in which $g(x)$ itself would differ from one observational cell to the next, and a fully parametric model in which $g(x)$ (as well as the cell effects) would be specified as a finitely parameterized function. The argument for this semiparametric strategy is analogous to the argument for specifying a parametric model with additive cell effects: Including the cell effects makes it possible to estimate benefits with cross-sectional data when there might be few observations in some cells. The specification of $g(x)$ imposes no parametric assumptions concerning the continuous part of the regression function.

When all of the observations are from the same cell, the proposed estimator is computed by first estimating the conditional expectation of Y_i, given X_i, using kernel nonparametric regression for each observation i in the sample. The conditional expectation of Y_i after the policy is implemented is estimated by evaluating the kernel regression estimator at X_i^* for each observation. The proposed estimator is the average difference between the kernel estimates of the conditional expectations at X_i^* and X_i. When there are cell-specific effects, this procedure is modified by subtracting estimates of these cell effects, which are first computed using a modification of least squares.

There has been much work on kernel nonparametric regression. Introduced by Nadaraya (1964) and Watson (1964), the pointwise consistency of kernel regression was proven by Devroye and Wagner (1980) and Spiegelman

* James H. Stock is Assistant Professor, John F. Kennedy School of Government, Harvard University, Cambridge MA 02138. This work was supported in part by National Science Foundation Grant SES-84-0879. The author thanks H. Bierens, J. Powell, D. Wise, two anonymous referees, and an associate editor for helpful suggestions on an earlier draft, and Rick White and Robin Lumsdaine for research assistance.

Journal of the American Statistical Association
June 1989, Vol. 84, No. 406, Theory and Methods

and Sacks (1980). Of particular importance in deriving the asymptotic properties of the proposed estimator are Devroye's (1978) and Bierens's (1983) uniform consistency results (under different conditions) for kernel regression. A variety of other theoretical results, including asymptotic normality (at a rate slower than $n^{1/2}$, where n is the sample size), were reviewed by Prakasa Rao (1983) and Bierens (1985). Alternative nonparametric regression techniques include spline regression (e.g., Wahba 1978), nearest-neighbor regression (e.g., Stone 1977), and flexible functional forms [e.g., the Fourier flexible functional forms of Gallant (1981) and Elbadawi, Gallant, and Souza (1983)]. Kernel regression is adopted here for computational convenience and because of existing theoretical results applicable to the benefits-estimation problem.

The semiparametric regression model and the proposed estimator are presented in Section 2. Asymptotic results concerning the behavior of the estimator are stated in Section 3. The estimator is shown to be consistent, and when centered around its conditional expectation, asymptotically normal. Furthermore, the estimator converges to this limiting distribution at $n^{1/2}$, faster than is typically exhibited by nonparametric estimators. This result complements Robinson's (1988) $n^{1/2}$ rate for the estimator of the parametric part of this semiparametric regression model. In Section 4, the properties of this estimator are investigated in a Monte Carlo study. This experiment confirms the theoretical prediction that, when the number of observations is small and the number of regressors large, the bias of the proposed estimator can be severe. This bias and its sources are discussed in Section 4. Conclusions are summarized in Section 5.

2. THE MODEL AND THE PROPOSED ESTIMATOR

2.1 The Model

The observations are assumed to be generated by a nonlinear version of the usual linear model, with m dummy variables. No parametric assumptions are made on the function relating the continuous independent variables to the dependent variable. The data consist of n observations from $m + 1$ observational cells on the dependent variable, Y_i, the k-dimensional continuous independent variables, X_i, and a vector of zeros and ones, d_i, indicating the cell from which the observation is drawn. Omitting one cell arbitrarily, the vector of dummy variables d_i has dimension m. With this notation, the model considered is

$$Y_i = g(X_i) + A'd_i + u_i, \qquad i = 1, \ldots, n, \qquad (1)$$

where A is the m-dimensional vector of cell-specific effects. It is assumed that (X_i, d_i, Y_i) are iid, $E(u_i \mid X_i, d_i) = 0$, and $E(u_i^2 \mid X_i, d_i) = \sigma^2(X_i, d_i)$, where there are constants $\underline{\sigma}^2$ and $\overline{\sigma}^2$ such that $0 < \underline{\sigma}^2 \le \sigma^2(x, d) \le \overline{\sigma}^2 < \infty$ for all (x, d) in the support of (X_i, d_i). The function $g(x)$ is assumed to be continuous in x.

I consider the problem of estimating the average change in the dependent variable Y resulting from a shift in the distribution of X. Let X_i^* denote the vector of policy and control variables after the intervention, and let $H(x)$ and $H^*(x)$ denote the marginal distributions of X_i and X_i^*. Before the intervention, the mean of Y_i is $EY_i = E[g(X_i) + A'd_i + u_i] = E[g(X_i) + A'd_i]$; after the intervention, this mean value is $E^*Y_i = E^*[g(X_i) + A'd_i]$, where $E^*[\cdot]$ denotes the expectation taken over X^*. The estimand is the mean effect on Y of the shift,

$$\begin{aligned} B &= E[g(X_i) + A'd_i] - E^*[g(X_i) + A'd_i] \\ &= E[g(X_i)] - E^*[g(X_i)], \qquad (2) \end{aligned}$$

where the second equality obtains by assuming that the policy does not alter the observational cell.

Note that (1) should be interpreted not simply as a conditional expectation, but as a structural model that is invariant to the proposed policy shift. Were this not so, the benefits expression (2) would not be valid, for $g(x)$ would differ before and after the policy is implemented. The possibility that $g(x)$ and A might change when the policy changes, perhaps because of sophisticated reactions of individuals to the government policy in question, has received considerable attention in the economics literature (this is often referred to as the "Lucas critique"). This possibility, however, is ruled out here by assumption. As a specific case in which the proposed estimator could be applied, consider the problem of predicting the average change in housing values that might arise from cleaning up a contaminated hazardous-waste disposal site. Let Y represent housing price, X represent a vector of housing attributes, including (for example) the distance to the nearest contaminated waste site, and $g(x)$ be a hedonic price equation describing the equilibrium relation between housing attributes and prices in a given metropolitan area. In this application, X denotes the housing attributes before cleaning up a given hazardous-waste site, X^* denotes the attributes after the cleanup, and B is the average cleanup benefit, measured in terms of increased housing prices. There has been considerable theoretical work on the use of hedonic housing-price surfaces to estimate the benefits of policy interventions, such as cleaning up a hazardous-waste site; for example, see Harrison and Rubinfeld (1978), Polinsky and Shavell (1975, 1976, 1978), and Scotchmer (1985). This work suggests that the assumption of an unchanged general equilibrium price equation is valid if the project is small relative to the total value of the houses in the metropolitan area (although the benefits need not be small for each house); it is unlikely to be valid if the project is large. More broadly, the assumption that $g(x)$ and A remain unchanged is conventional (if controversial) in parametric policy analyses, and it is maintained in the nonparametric treatment here as well.

2.2 The Proposed Estimator

The proposed estimator sidesteps the problem of specifying a functional form for $g(x)$ through the use of kernel nonparametric regression. The basic idea of the estimator is a simple one: For the ith observation, obtain consistent nonparametric estimates of $E(Y \mid X_i)$ and $E(Y \mid X_i^*)$. Repeat this for each observation, $i = 1, \ldots, n$. The difference between the estimated conditional expectations at X_i^* and X_i provides an estimate of the effect of the proposed policy shift on Y_i for each i. The estimator of the mean benefits is the average of each of these individually estimated benefits.

When there are no dummy variables, the estimator involves direct averages of nonparametric estimates of the regression function. Let the kernel weight function $w(t)$ be a density on $\Re^k$ (technical conditions are given in Sec. 3), b_n be the kernel bandwidth parameter, and $g_n(x)$ denote the kernel estimator of $g(x)$:

$$g_n(x) = \sum_{i=1}^{n} w((X_i - x)/b_n)Y_i \Big/ \sum_{i=1}^{n} w((X_i - x)/b_n). \tag{3}$$

Under weak conditions on the densities, if $b_n \to 0$ and $nb_n^k \to \infty$, then $g_n(x)$ is a consistent estimator of $g(x)$ (Spiegelman and Sacks 1980). The proposed estimator is simply the sample analog of (2), computed using the kernel estimator (3) evaluated at all sample points:

$$B_n = n^{-1} \sum_{j=1}^{n} [g_n(X_j^*) - g_n(X_j)], \tag{4}$$

where $g_n(X_j^*)$ and $g_n(X_j)$ are the kernel estimators of $g(X_j^*)$ and $g(X_j)$, respectively, both constructed using (3). [Note that one could alternatively estimate B by $\tilde{B}_n = n^{-1} \sum_{j=1}^{n} [g_n(X_j^*) - Y_j]$. I focus on B_n, although the theoretical treatment of these two estimators is similar.]

The presence of dummy variables complicates the problem considerably, since it is now necessary to estimate the nuisance parameter A in (1) as well as the mean benefits. To motivate the proposed technique, recall the ordinary least squares (OLS) estimator of A when $g(x)$ is linear. Adopting the usual matrix notation, the linear version of (1) is

$$Y = X\beta + DA + U. \tag{5}$$

The parameters β and A in (5) can be estimated using OLS with both X and D as right-hand variables simultaneously. Alternatively, the OLS estimator of A can be written as $\hat{A} = (D'M_XD)^{-1}D'M_XY$, where $M_X = I - X(X'X)^{-1}X'$. That is, $\hat{A}$ can be computed by regressing the residuals of a regression of Y on X against the residuals of a regression of D on X.

When $g(x)$ is unknown, there is no clear way to estimate g and A simultaneously. Instead, A, g, and B are estimated in three steps. The first step is the estimation of A, which is motivated by analogy to the OLS estimator: Estimate the cell effects using OLS by regressing the residuals from a kernel nonparametric regression of Y on X, against the residuals from a kernel regression of d on X. Let $f_1(x)$ and $f_2(x)$ (respectively) denote the conditional expectations of Y and d, given x, and let $f_{1n}(x)$ and $f_{2n}(x)$ denote their respective kernel estimators:

$$f_1(x) \equiv E(Y \mid x), \tag{6}$$

$$f_2(x) \equiv E(d \mid x), \tag{7}$$

$$f_{1n}(x) = \sum_{i=1}^{n} w((X_i - x)/b_n)Y_i \Big/ \sum_{i=1}^{n} w((X_i - x)/b_n), \tag{8}$$

and

$$f_{2n}(x) = \sum_{i=1}^{n} w((X_i - x)/b_n)d_i \Big/ \sum_{i=1}^{n} w((X_i - x)/b_n). \tag{9}$$

Let η_i and ξ_i denote the residuals from these kernel regressions,

$$\eta_i = Y_i - f_{1n}(X_i) \tag{10}$$

and

$$\xi_i = d_i - f_{2n}(X_i). \tag{11}$$

A is then estimated by the OLS regression of η_i onto ξ_i:

$$A_n = \left(\sum_{j=1}^{n} \xi_j \xi_j'\right)^{-1} \left(\sum_{j=1}^{n} \xi_j \eta_j\right). \tag{12}$$

It is shown in the next section that A_n is consistent for A.

The second step involves obtaining a consistent nonparametric estimator of $g(x)$, given this consistent estimator of A. Such an estimator can be obtained by noting that (1) and (7) imply that $E(Y \mid x) = g(x) + E(d \mid x)'A = g(x) + f_2(x)'A$, whereas (6) states that $E(Y \mid x) = f_1(x)$. Thus

$$g(x) = f_1(x) - f_2(x)'A. \tag{13}$$

Although $g(x)$ cannot be estimated directly, each component on the right side of (13) can. Accordingly, $g(x)$ can be estimated by

$$g_n(x) = f_{1n}(x) - f_{2n}(x)'A_n. \tag{14}$$

The third step in computing the benefits estimator is to evaluate $g_n(x)$ at each sample value of X and X^*. When there are cell effects, as long as the policy does not change the cell in which the observation is located the estimator still has the form (4), with the regression estimator in (14) replacing the simpler one in (3).

Combining the various expressions for B_n and $g_n(x)$, the proposed estimator is

$$B_n = n^{-1} \sum_{j=1}^{n} \gamma_n(X_j)(Y_j - d_j'A_n), \tag{15}$$

where

$$\gamma_n(x) = \lambda_n^*(x) - \lambda_n(x),$$

$$\lambda_n^*(x) = \sum_{i=1}^{n} \left[w((x - X_i^*)/b_n) \Big/ \sum_{j=1}^{n} w((X_j - X_i^*)/b_n) \right],$$

and

$$\lambda_n(x) = \sum_{i=1}^{n} \left[w((x - X_i)/b_n) \Big/ \sum_{j=1}^{n} w((X_j - X_i)/b_n) \right],$$

where A_n is given in (12). Note that both the OLS and the nonparametric benefits estimators are linear in the dependent variable, with weights that depend solely on $\{X_i, X_i^*, d_i\}$.

3. CONSISTENCY AND ASYMPTOTIC NORMALITY

This section presents asymptotic results for the cell-effects estimator A_n and the benefits estimator B_n. Both A_n and B_n are consistent. In addition, when centered around its expectation conditional on $\{X_i, d_i\}$ $(i = 1, \ldots, n)$, the benefits estimator is asymptotically normal.

The following assumptions are made concerning the distributions of X and X^* and the conditional expectations $E(Y \mid X)$ and $E(d \mid X)$.

Assumption 1. (a) $H(x)$ and $H^*(X)$ (respectively) have continuous densities $h(x)$ and $h^*(x)$. In addition, H and H^* have a common compact support Ξ, and $\exists\, h_1$ and h_2 such that $0 < h_1 \le h(x), h^*(x) \le h_2 < \infty$ for all $x \in \Xi$. (b) $f_1(x)$ and $f_2(x)$ are bounded and continuous in x uniformly over Ξ. (c) $0 < \int [h^*(x)/h(x) - 1]^2\, dH(x) < \infty$.

The assumption that X and X^* have the same support is not innocuous: It restricts the policy experiments that can be considered to ones for which there already exists some experience in the data. This is a consequence of the inapplicability of kernel regression to extrapolation.

The kernel $w(u)$ is assumed to satisfy the following condition.

Assumption 2. $w(u)$ is a symmetric, everywhere-positive density on $\Re^k$ with an absolutely integrable characteristic function.

Define

$$R_n = n^{-1} \sum_{j=1}^{n} \gamma_n(X_j) d_j,$$

$$R = \int E(d \mid x)\, dH^*(x) - \int E(d \mid x)\, dH(x),$$

$$M_n = n^{-1} \sum_{j=1}^{n} \xi_j \xi_j',$$

$$M = E[(d_i - E(d_i \mid X_i))(d_i - E(d_i \mid X_i))'],$$

and

$$\gamma(x) = h^*(x)/h(x) - 1.$$

The first two results are that the cell-effects estimator A_n and the benefits estimator B_n are consistent.

Theorem 1. If Assumptions 1 and 2 hold, $b_n \to 0$, and $nb_n^{2k} \to \infty$, then $M_n \xrightarrow{p} M$ and $A_n \xrightarrow{p} A$.

Theorem 2. Under the conditions of Theorem 1, $B_n \xrightarrow{p} B$.

The proofs of these theorems are given in the Appendix.

The rates of convergence to 0 of the bandwidths in Theorems 1 and 2 are slower than needed for pointwise consistency, which requires that $nb_n^k \to \infty$ rather than $nb_n^{2k} \to \infty$. This slower rate is one of Bierens's (1983) conditions for the uniform consistency of kernel regression, a result used to prove Theorems 1 and 2.

The consistency of B_n arises from the uniform consistency of the weights $\gamma_n(x)$ or equivalently the uniform consistency of the kernel estimators of $f_1(x)$ and $f_2(x)$. Although the kernel regression estimator is consistent at each point in Ξ, it converges to its pointwise probability limit at a rate slower than $n^{1/2}$. This difficulty also arises with the nonparametric benefits estimator: If $g(x)$ is nonlinear, then $E[B_n \mid \{X_i, d_i\}]$ converges to B at a rate slower than $n^{1/2}$.

Despite the slow rate of convergence of the conditional mean of B_n to B, when centered at $E[B_n \mid \{X_i, d_i\}]$ the estimator is asymptotically normal and converges to its limiting distribution at the rate $n^{1/2}$.

Theorem 3. Suppose that Assumptions 1 and 2 hold, $b_n \to 0$, $nb_n^{2k} \to \infty$, and that there are constants $\delta > 0$ and Δ such that $E|u_j|^{2+\delta} < \Delta < \infty$ for all j. Then, $n^{1/2}(B_n - E[B_n \mid \{X_i, d_i\}]) \xrightarrow{\mathcal{L}} N(O, V)$, where $V = E[\sigma^2(X, d)(\gamma(X) - R'M^{-1}[d - f_2(X)])^2]$. In addition, $V_n = n^{-1}\sum_{j=1}^{n} (\gamma_n(X_j) - R_n' M_n^{-1} \pi_{nj})^2 u_{nj}^2 \xrightarrow{p} V$, where $\pi_{nj} = \xi_j - \sum_{i=1}^{n} w((X_j - X_i)/b_n)\xi_i / \sum_{i=1}^{n} w((X_j - X_i)/b_n)$ and $u_{nj} = Y_j - g_n(X_j) - A_n' d_j$. If $\sigma^2(x, d) = \sigma^2$ for all (x, d), then $V = \sigma^2[\int (h^*(x)/h(x) - 1)^2\, dH(x) + R'M^{-1}R]$. The proof is given in the Appendix.

4. MONTE CARLO RESULTS

This section presents the results of Monte Carlo simulations performed for models with one policy variable, X_1, and from zero to two control variables. The data were generated by a linear version of (1):

$$Y_{ij} = 1 + \sum_{r=1}^{k} X_{rij} + A_j + u_{ij}, \qquad (16)$$

for $j = 1, \ldots, m$. Each observation on X_{rij} was drawn independently from a uniform distribution on the unit interval. The errors u_j were drawn from an $NI(0, .25)$ distribution. The simulated shift in the policy variable was $X_{1ij}^* = X_{1ij}^2$ for all i and j. Under these assumptions, the true value of B is $-.1667$. When the observations were drawn from more than one cell, $[n/(m + 1)]$ observations were drawn from the first m cells and the remaining observations were drawn from the final cell, where $[\cdot]$ denotes the greatest lesser integer. The estimator was computed using a multivariate Gaussian kernel, where the sample covariance matrix of X was used as the covariance matrix in the kernel. The bandwidths were computed using $b_n = (b/n^{1/2})^{1/k}$, where b is a parameter varied across simulations.

This study examines B_n and two variance estimators suggested by the expressions for V_n given in Theorem 3. The sum of squared residuals from the usual kernel regression can understate the regression variance in small samples, since Y_i and d_i enter $f_{1n}(X_i)$ and $f_{2n}(X_i)$ with relatively large weights. Thus the variances were estimated using the residuals $\hat{u}_{nj}$ from the "drop-j" kernel regressions,

$$\hat{u}_{nj} = Y_j - [f_{1n(j)}(X_j) + (d_j - f_{2n(j)}(X_j))' A_n], \qquad (17)$$

where

$$f_{1n(j)}(X_j) = \sum_{i \ne j} w((X_i - X_j)/b_n) Y_i \Big/ \sum_{i \ne j} w((X_i - X_j)/b_n) \qquad (18)$$

and

$$f_{2n(j)}(X_j) = \sum_{i \neq j} w((X_i - X_j)/b_n)d_i \Big/ \sum_{i \neq j} w((X_i - X_j)/b_n) \tag{19}$$

(e.g., see Li 1984; Devroye and Penrod 1984; Marron 1985; Rice 1984a). Using these residuals, the variance estimators considered are

$$V_{1n} = n^{-1} \sum_{i=1}^{n} c_{ni}^2 \hat{u}_{ni}^2 \tag{20}$$

and

$$V_{2n} = \left(n^{-1} \sum_{i=1}^{n} c_{ni}^2\right)\left(n^{-1} \sum_{i=1}^{n} \hat{u}_{ni}^2\right), \tag{21}$$

where $c_{ni} = \gamma_n(X_i) - R_n' M_n^{-1} \pi_{ni}$. The simulations were computed using 100 draws for $n = 20$, 40, and 60, and 50 draws for $n = 100$.

Selected simulation results are presented in Table 1. Several features are apparent from these results. Even with $n = 40$, the bias discussed in the preceding section can be small when $k = 1$. As k increases, however, the estimator is increasingly biased toward 0. With large k, this bias can be substantial, even for large n. The bias of the estimator also grows as the number of cells increases, although this deterioration does not seem to be as important as that associated with having more control variables. The effect of m is most pronounced for small sample sizes; for example, for $n = 60$ and $k = 3$ the mean of B_n changes only slightly as m increases. The two variance estimators have similar means, particularly as n increases. In the larger samples the averages of the variance estimators generally fall within 10% of the simulation variance and typically are conservative (as is expected using the drop-j regression residuals).

As an additional comparison, Table 1 presents the ratio of the root mean square error (RMSE) of the correctly specified OLS estimator of B [which is efficient under (16)] to the RMSE of B_n. In all cases this ratio exceeds 50%, and in some it exceeds 80%. The comparison is most favorable to B_n when its bias is smallest. Thus for this model the cost of using the nonparametric estimator appears to be modest.

Three distinct sources of the bias in B_n were investigated using response-surface regressions. The first arises from Jensen's inequality when estimating $g_n(x)$ for x in the in-

Table 1. Monte Carlo Results Based on Model (13)

n	k	m	$\bar{B}_n$	$var(B_n) \times 10^{-2}$	$\bar{V}_1/n \times 10^{-2}$	$\bar{V}_2/n \times 10^{-2}$	$\bar{A}_1$	$\bar{A}_2$	$\bar{A}_3$	$\lvert(\bar{B}_n - B)/B\rvert$	$RMSE(\hat{B})/RMSE(B_n)$
40	1	0	−.171	.474	.763	.695	—	—	—	.027	.721
40		1	−.148	.712	.869	.651	.785	—	—	.114	.584
40		2	−.169	.705	.692	.670	.806	−.782	—	.013	.601
40		3	−.149	.707	.873	.832	.798	−.758	1.564	.106	.598
40	2	0	−.130	.301	.325	.342	—	—	—	.220	.767
40		1	−.119	.283	.337	.328	.676	—	—	.286	.707
40		2	−.126	.324	.376	.363	.679	−.698	—	.245	.738
40		3	−.126	.329	.503	.469	.678	−.686	1.360	.244	.749
40	3	0	−.096	.170	.203	.198	—	—	—	.425	.632
40		1	−.100	.198	.214	.223	.554	—	—	.401	.644
40		2	−.098	.247	.288	.300	.561	−.555	—	.414	.615
40		3	−.093	.276	.357	.375	.543	−.553	1.080	.443	.578
60	1	0	−.165	.380	.442	.447	—	—	—	.012	.639
60		1	−.161	.365	.403	.423	.833	—	—	.031	.658
60		2	−.163	.474	.451	.452	.823	−.830	—	.024	.581
60		3	−.168	.395	.598	.467	.815	−.806	1.657	.008	.652
60	2	0	−.144	.198	.222	.215	—	—	—	.134	.801
60		1	−.129	.176	.231	.220	.734	—	—	.223	.723
60		2	−.123	.227	.283	.264	.721	−.730	—	.263	.631
60		3	−.138	.312	.322	.313	.768	−.731	1.476	.175	.651
60	3	0	−.112	.110	.134	.139	—	—	—	.327	.633
60		1	−.112	.165	.169	.160	.642	—	—	.328	.601
60		2	−.113	.194	.194	.198	.621	−.626	—	.322	.595
60		3	−.113	.217	.252	.248	.612	−.619	1.234	.320	.574
100	1	0	−.171	.244	.260	.262	—	—	—	.027	.612
100		1	−.153	.318	.231	.238	.864	—	—	.082	.525
100		2	−.171	.232	.281	.267	.869	−.847	—	.028	.630
100		3	−.171	.189	.266	.280	.861	−.878	1.744	.025	.706
100	2	0	−.144	.148	.139	.137	—	—	—	.136	.682
100		1	−.150	.115	.141	.140	.774	—	—	.103	.806
100		2	−.151	.206	.153	.150	.809	−.782	—	.096	.637
100		3	−.143	.125	.162	.165	.788	−.805	1.563	.143	.727
100	3	0	−.123	.066	.091	.091	—	—	—	.264	.599
100		1	−.123	.089	.104	.101	.708	—	—	.261	.571
100		2	−.128	.109	.064	.111	.696	−.686	—	.232	.570
100		3	−.129	.164	.147	.134	.690	−.705	1.377	.225	.564

NOTE: The true values of the cell effects are $A_1 = 1$, $A_2 = -1$, and $A_3 = 2$. The bandwidth was computed using $b = 1$. The results for $n = 40$ and 60 were produced using 100 replications; 50 replications were used for $n = 100$. The parameters describing the simulated model are given in the first three columns. The average of the simulated estimates B_n is given in the fourth column, and the fifth column contains the variance estimated from the Monte Carlo sample. The next two columns provide the averages of V_{1n} and V_{2n}, and the next three columns show the averages of A_{1n}, A_{2n}, and A_{3n}. The second-to-last column presents the absolute relative bias of the nonparametric estimator. The final column presents the ratio of the RMSE of the OLS estimator of B to the RMSE of the nonparametric estimator.

terior of the support of X; this occurs when either the regression function $g(x)$ or the mapping from X to X^* is nonlinear. The bias at an interior point is proportional to the square of the bandwidth. Integrating (with respect to H) over the support of X, the contribution of this source of bias to the total bias of the estimator has a Taylor series expansion in which the leading term is approximately $b_n^2(1 - \rho)$, where ρ is the (H) measure of the support of X within a bandwidth of its boundary. (This leading term would be exact if an indicator function with bandwidth b_n were used as the kernel, and is approximate for the Gaussian kernel.)

The second source of bias arises for observations on X or X^* that are close to the boundary of the support of X; in this case, the bias generally occurs when the first derivative of either the regression function or the policy shift is nonzero in the neighborhood of the boundary. This bias can be expressed as a Taylor series expansion with terms $b_n\rho$, $b_n^2\rho$, and so forth. Unlike the bias for points in the interior, in general the coefficients on all terms in this expansion are nonzero. The third source of bias arises from estimating the cell effects. Absent a theoretical specification for the rate at which this bias decreases, it is modeled by m/n.

Estimated response-surface regressions for the bias are presented in Table 2. The data for these response surfaces were generated by Monte Carlo simulations of B_n using the 144 combinations of the parameter values $n = 20, 40, 60, 100$; $k = 1, 2, 3$; $m = 0, 1, 2, 3$; and $b = 1, 2, 3$. Terms in k/n and $1/n$ were included in the regressions, in addition to those discussed previously. Since the right-hand variables have limits of 0 as n increases to infinity, an intercept of 0 in these regressions indicates 0 asymptotic bias. In all regressions except that using the $m = 1$ subset of the simulated data, the intercept term is statistically indistinguishable from 0 at the 10% level; in the $m = 1$ regression the intercept, although significantly different from 0 at the 5% level, is numerically small (less than 4% of B). Thus the response-surface regressions reflect the consistency of B_n.

The regressions in Table 2 suggest that the primary source of finite-sample bias is the contribution of observations on the boundary of the support of X. For example, for $n = 60$, $m = 2$, $k = 2$, and $b_n = .25$, the two terms corresponding to the boundary bias ($b_n\rho$ and $b_n^2\rho$) contribute .040, whereas the other terms contribute .009. The results also suggest that the bias is not greatly affected by increasing the number of cells, m; even for $m = 2$ and $n = 20$, the contribution of this term (m/n) to the bias is only .003.

In summary, this experiment suggests five conclusions. First, the most important source of the bias of the proposed estimator appears to be boundary effects. Second, the bias of the estimator is not substantially affected by increasing the number of cells from which the observations are drawn. Third, as expected, the number of explanatory variables is an important factor contributing to the bias. Fourth, even accounting for this bias, the estimator has reasonable performance relative to the efficient estimator (OLS) in this design. Fifth, even when n is small, the proposed variance estimators perform well (in the sense that their bias is small), at least for the models studied.

These results emphasize the importance of developing a $n^{1/2}$-consistent version of the estimator. One approach to the interior bias problem involving jackknifed nonparametric regression estimators was developed by Bierens (1985) and applied by Powell, Stock, and Stoker (in press). Alternatively, it might be possible to handle this source of bias using higher-order kernels (e.g., Hall and Marron 1988). For an extrapolation-based solution to the boundary bias problem with $k = 1$ and $m = 0$, see Rice (1984b).

5. CONCLUSIONS

The estimator introduced in this article provides a way to avoid one of the ambiguities associated with predicting the effects of a proposed policy, specifically the choice of a parametric regression function. The proposed estimator is consistent and, when centered around its conditional expectation, asymptotically normal under general conditions on the continuous part of the regression function. In a Monte Carlo experiment, asymptotic variance expressions were found to provide an approximate guide to the sampling variance, even in samples of only 40 observations. The nonparametric estimator incurred only a moderate increase in RMSE relative to the efficient parametric estimator (OLS). Nevertheless, these simulations point to two areas warranting further work: bandwidth selection

Table 2. Estimator Bias: Response-Surface Regressions

Subset	Constant	$b_n^2(1-\rho)$	$b_n\rho$	$b_n^2\rho$	m/n	k/n	$1/n$	No. of observations	R^2
All	−.002	.084	.289	−.298	.028	.344	−.182	144	.948
	(.002)	(.012)	(.029)	(.029)	(.015)	(.044)	(.098)		
$m = 0$	−.003	.085	.285	−.303	—	.433	−.314	36	.966
	(.003)	(.021)	(.052)	(.051)		(.078)	(.170)		
$m = 1$	.006	.070	.224	−.202	—	.273	−.147	36	.968
	(.002)	(.018)	(.043)	(.042)		(.064)	(.141)		
$m = 2$	−.004	.074	.344	−.356	—	.269	.123	36	.958
	(.003)	(.024)	(.059)	(.057)		(.087)	(.189)		
$m = 3$	−.005	.108	.300	−.330	—	.408	−.235	36	.943
	(.004)	(.029)	(.071)	(.069)		(.107)	(.232)		

NOTE: Standard errors are in parentheses. The data set consists of the simulations discussed in the text.

and the possibility of reducing the size of the bias. The response-surface regressions indicate that the primary contribution to the bias seems to come from observations near the boundary rather than in the interior. This suggests that small-sample performance might be improved using variable-size bandwidths or kernels that can take on negative values.

In a separate application (Stock, in press), this estimator was applied to data on attributes and sales prices of single-family homes in the Boston, Massachusetts, area. The data included information about local hazardous-waste sites, and the policy experiment simulated was cleaning up one of the 11 hazardous-waste sites in the study area, holding other house attributes constant. The procedure suggested that the aggregate benefits of cleaning up a site, in terms of increased house prices, could be substantial compared to engineering estimates of the cost. The major methodological conclusion from this investigation was that additional theoretical guidance concerning the problems of kernel and bandwidth selection for this estimator could prove to be of considerable practical value.

APPENDIX: PROOFS OF THEOREMS

Theorem 1. The proof uses Bierens's (1983) uniform consistency result for kernel regression. Let $f_n(x) = (f_{1n}(x)\ f_{2n}(x)')'$ and $f(X) = (f_1(X)\ f_2(X)')'$. Under weaker conditions than Assumptions 1 and 2, Bierens proved that if $nb_n^{2k} \to \infty$,

$$\sup_x |f_n(x) - f(x)| \xrightarrow{p} 0, \tag{A.1}$$

where the supremum is taken over the compact support Ξ of X and the convergence applies to each element of the vectors $f_n(X)$ and $f(x)$.

Turning to the proof, from (12), $A_n = (n^{-1}\Sigma_j \xi_j \xi_j')^{-1}(n^{-1}\Sigma_j \xi_j \eta_j)$. Using (10) and (11), one obtains

$$n^{-1}\sum_j \xi_j \xi_j' = n^{-1}\sum_j (d_j - f_{2n}(X_j))(d_j - f_{2n}(X_j))'$$
$$= M_n^* + J_{1n} + J_{1n}' + J_{2n}, \tag{A.2}$$

where

$$M_n^* = n^{-1}\sum_j (d_j - f_2(X_j))(d_j - f_2(X_j))',$$
$$J_{1n} = n^{-1}\sum_j (d_j - f_2(X_j))(f_2(X_j) - f_{2n}(X_j))',$$

and

$$J_{2n} = n^{-1}\sum_j (f_2(X_j) - f_{2n}(X_j))(f_2(X_j) - f_{2n}(X_j))'.$$

Similarly, using (10) and (11) and recalling from (12) and (13) that $Y_j = f_1(X_j) + (d_j - f_2(X_j))'A + u_j$, one obtains

$$n^{-1}\sum_j \xi_j \eta_j = (M_n^* + J_{1n}')A + J_{3n} + J_{4n} + J_{5n} + J_{6n}, \tag{A.3}$$

where

$$J_{3n} = n^{-1}\sum_j (d_j - f_2(X_j))(f_1(X_j) - f_{1n}(X_j)),$$
$$J_{4n} = n^{-1}\sum_j (f_2(X_j) - f_{2n}(X_j))(f_1(X_j) - f_{1n}(X_j)),$$
$$J_{5n} = n^{-1}\sum_j (d_j - f_2(X_j))u_j,$$

and

$$J_{6n} = n^{-1}\sum_j (f_2(X_j) - f_{2n}(X_j))u_j.$$

It is now shown that $J_{in} \xrightarrow{p} 0$ ($i = 1, \ldots, 6$), and that $M_n^* \xrightarrow{p} M$; from (A.2) and (A.3), and the assumption that M is positive definite (so that M^{-1} exists), it follows that $M_n \xrightarrow{p} M$ and $A_n \xrightarrow{p} A$. Turning to the J_{in} terms, note that element by element $J_{1n} \le 2(\max_j |d_j|)\ (\sup_x |f_{2n}(x) - f_2(x)|)$. Since d_j is a vector of zeros and ones, the uniform consistency result (A.1) implies that $J_{1n} \xrightarrow{p} 0$. Similarly, element by element $|J_{2n}| \le (\sup_x |f_{2n}(x) - f_2(x)|)^2 \xrightarrow{p} 0$, by (A.1). The argument that $J_{1n} \xrightarrow{p} 0$ applies directly to J_{3n} and the argument used for J_{2n} applies to J_{4n}. Because (X_j, d_j, u_j) is independent of (X_i, d_i, u_i) $(i \ne j)$, $J_{5n} \xrightarrow{p} 0$ by the weak law of large numbers (WLLN). Considering J_{6n}, element by element $|J_{6n}| \le (\sup_x |f_2(x) - f_{2n}(x)|)(n^{-1}\Sigma_j |u_j|)$. The independence of u_j [with $Eu_j^2 \le (\bar{\sigma}^2)$] and the uniform consistency of $f_{2n}(x)$ imply that $J_{6n} \xrightarrow{p} 0$. Finally, because $EM_n^* = M$ and because (d_j, X_j) are bounded and iid, $M_n^* \xrightarrow{p} M$. Thus $M_n \xrightarrow{p} M$ and $A_n \xrightarrow{p} A$.

Theorem 2. A direct proof obtains using Theorem 1 and the uniform consistency of $f_{1n}(x)$ and $f_{2n}(x)$. The proof given here, however, is developed for the expression (15), and establishes some results that are useful in proving Theorem 3. Let $B_n^* = n^{-1}\Sigma_j \gamma_n(X_j)(Y_j - d_j'A)$ so that $B_n = B_n^* - n^{-1}\Sigma_j \gamma_n(X_j)d_j'(A_n - A) = B_n^* - R_n'(A_n - A)$. It is shown that (a) $R_n \xrightarrow{p} R < \infty$ and (b) $B_n^* \xrightarrow{p} B$. Given (a) and (b), it follows from Theorem 1 that $B_n \xrightarrow{p} B$.

(a) Using the definitions of R_n, $\gamma_n(x)$, $\lambda_n^*(x)$, and $\lambda_n(x)$,

$$R_n = n^{-1}\sum_j \lambda_n^*(X_j)d_j - n^{-1}\sum_j \lambda_n(X_j)d_j$$
$$= n^{-1}\sum_j \left[\sum_i w((X_i^* - X_j)/b_n)d_j \Big/ \sum_j w((X_i^* - X_j)/b_n)\right]$$
$$- n^{-1}\sum_j \left[\sum_i w((X_i - X_j)/b_n)d_j \Big/ \sum_j w((X_i - X_j)/b_n)\right]$$
$$= n^{-1}\sum_i f_{2n}(X_i^*) - n^{-1}\sum_i f_{2n}(X_i).$$

Now,

$$n^{-1}\sum_i f_{2n}(X_i^*) - \int f_2(x)\, dH^*(X)$$
$$= n^{-1}\sum_i [f_{2n}(X_i^*) - f_2(X_i^*)]$$
$$+ \left[n^{-1}\sum_i f_2(X_i^*) - \int f_2(x)\, dH^*(x)\right]. \tag{A.4}$$

The first term on the right side of (A.4) vanishes by the uniform consistency of $f_{2n}(x)$, and since $f_2(x)$ is bounded the second set of terms in (A.4) is $o_p(1)$ using the WLLN. Thus $n^{-1}\Sigma_i f_{2n}(X_i^*) \xrightarrow{p} E^*[f_2(x)] = \int E(d \mid x)\, dH^*(x)$. The same argument applies to $n^{-1}\Sigma_i f_{2n}(X_i)$, so $R_n \xrightarrow{p} \int E(d \mid x)[h^*(x) - h(x)]\, dx = R$.

(b) Before showing that $B_n^* \xrightarrow{p} B$, it is first shown that $\sup_x |\gamma_n(x) - \gamma(x)| \xrightarrow{p} 0$. From its definition,

$$\lambda_n^*(x) = \sum_i w((X_i^* - x)/b_n) \Big/ \sum_j w((X_i^* - X_j)/b_n)$$
$$= (nb_n^k)^{-1}\sum_i w((X_i^* - x)/b_n)/h(X_i^*)$$
$$+ (nb_n^k)^{-1}\sum_i w((X_i^* - x)/b_n)(h_n(X_i^*)^{-1} - h(X_i^*)^{-1}), \tag{A.5}$$

where $h_n(x) = (nb_n^k)^{-1}\Sigma_i w((X_i - x)/b_n)$ is the kernel density estimator of $h(x)$. Using $w(u) \ge 0$, the second term in the final expression in (A.5) vanishes asymptotically:

$$(nb_n^k)^{-1} \sum_i w((X_i^* - x)/b_n)(h_n(X_i^*)^{-1} - h(X_i^*)^{-1})$$
$$\le (nb_n^k)^{-1} \sum_i w((X_i^* - x)/b_n)\sup_x|h_n(x)^{-1} - h(x)^{-1}|$$
$$\le (\sup_x|h_n^*(x)|)(\sup_x|h_n(x) - h(x)|)/\inf_x|h(x)h_n(x)|,$$

where $h_n^*(x) = (nb_n^k)^{-1} \sum_i w((X_i^* - x)/b_n)$. Since $h_n(x) \xrightarrow{p} h(x)$ uniformly, and by Assumption 1(a), $0 < h_1 \le h(x) \le h_2 < \infty$, the bound in the final inequality converges in probability to 0. Turning to the first term of the final expression in (A.5), since $h(x)$ is continuous, $(nb_n^k)^{-1} \sum_i w((X_i^* - x)/b_n)/h(X_i^*) \xrightarrow{p} h^*(x)/h(x)$ uniformly in x. Thus $\lambda_n^*(x) \xrightarrow{p} \lambda^*(x) \equiv h^*(x)/h(x)$ uniformly in x. Applying the same argument to $\lambda_n(x)$, $\lambda_n(x) \xrightarrow{p} 1$ uniformly in x, from which it follows that $\sup_x|\gamma_n(x) - \gamma(x)| \xrightarrow{p} 0$.

Turning to B_n^*, use (1) to write

$$B_n^* = n^{-1} \sum_j \gamma_n(X_j)(Y_j - d_j'A)$$
$$= n^{-1} \sum_j \gamma_n(X_j)g(X_j) + n^{-1} \sum_j \gamma_n(X_j)u_j.$$

Since $\gamma(x)$ and $g(x)$ are bounded on Ξ [the latter follows from Assump. 1(b)], the uniform consistency of $\gamma_n(x)$ and the WLLN imply that

$$n^{-1} \sum_j \gamma_n(X_j)g(X_j) = n^{-1} \sum_j \gamma(X_j)g(X_j) + o_p(1)$$
$$\xrightarrow{p} \int g(x)[h^*(x) - h(x)]\, dx = B.$$

In addition,

$$E\left[\left(n^{-1} \sum_j (\gamma_n(X_j) - \gamma(X_j))u_j\right)^2 \middle| \{X_i, d_i\}\right]$$
$$\le n^{-1} (\sup_x|\gamma_n(X_j) - \gamma(X_j)|)^2\bar{\sigma}^2 \xrightarrow{p} 0,$$

and $n^{-1} \sum_j (\gamma_n(X_j) - \gamma(X_j))u_j \xrightarrow{p} 0$. Thus $n^{-1} \sum_j \gamma_n(X_j)u_j = n^{-1} \sum_j \gamma(X_j)u_j + o_p(1) \xrightarrow{p} 0$ by the boundedness of $\gamma(x)$ and the WLLN. Thus $B_n^* \xrightarrow{p} B$, implying that $B_n \xrightarrow{p} B$.

Theorem 3. The proof has two steps. First, it is shown that

$$n^{1/2}\left[(B_n - E(B_n \mid \{X_i, d_i\})) - n^{-1} \sum_j c_j u_j\right] \xrightarrow{p} 0, \quad \text{(A.6)}$$

where $c_j \equiv c(X_j, d_j) = \gamma(X_j) - R'M^{-1}\pi_j$, where $\pi_j = d_j - f_2(X_j)$. It follows from (A.6) that the centered benefits estimator and $n^{-1/2} \sum c_j u_j$ have the same limiting distribution. Second, it is shown that the independent random variables $Z_j = c_j u_j$ satisfy the conditions of Lyapunov's central limit theorem, with the variance V stated in Theorem 3.

To proceed with the first step, define

$$c_n(X_j, d_j) = \gamma_n(X_j) - R_n'M_n^{-1}\pi_{nj}$$

and

$$W_{ij} = w((X_i - X_j)/b_n)\Big/\sum_i w((X_i - X_j)/b_n),$$

where π_{nj} is defined in the theorem. With this notation, $\eta_j = Y_j - f_{1n}(X_j) = Y_j - \sum_i W_{ij}Y_i$. Thus, from (12),

$$A_n = M_n^{-1}n^{-1} \sum_j \xi_j\left(Y_j - \sum_i W_{ij}Y_i\right)$$
$$= M_n^{-1}n^{-1} \sum_j \left(\xi_j - \sum_k W_{jk}\xi_k\right) Y_j = M_n^{-1}n^{-1} \sum_j \pi_{nj}Y_j.$$

Using (15) and the definitions of R_n and $c_n(X_j, d_j)$, one obtains

$$B_n = n^{-1} \sum_j \gamma_n(X_j)(Y_j - d_j'A_n)$$
$$= n^{-1} \sum_j \gamma_n(X_j)\left(Y_j - d_j'M_n^{-1}n^{-1} \sum_i \pi_{ni}Y_i\right)$$
$$= n^{-1} \sum_j c_n(X_j, d_j)Y_j.$$

Note that $c_n(X_j, d_j)$ is a function solely of $\{(X_i, d_i)\}$, and use (1) to obtain

$$n^{1/2}[B_n - E(B_n \mid \{X_i, d_i\})]$$
$$= n^{-1/2} \sum_j c_n(X_j, d_j)u_j$$
$$= n^{-1/2} \sum_j [c_n(X_j, d_j) - c(X_j, d_j)]u_j$$
$$+ n^{-1/2} \sum_j c(X_j, d_j)u_j.$$

Turning to the first term in (A.6), if $c_n(x, d) - c(x, d) \xrightarrow{p} 0$ uniformly in (x, d), then with probability 1,

$$E\left[\left(n^{-1/2} \sum_j (c_n(X_j, d_j) - c(X_j, d_j))u_j\right)^2 \middle| \{x_i, d_i\}\right]$$
$$\le (\sup_{(x,d)}|c_n(x, d) - c(x, d)|)^2\bar{\sigma}^2 \xrightarrow{p} 0. \quad \text{(A.7)}$$

It follows from (A.7) and Chebyschev's inequality (modified for conditional expectations) that $n^{-1/2} \sum_j (c_n(X_j, d_j) - c(X_j, d_j))u_j \xrightarrow{p} 0$, if it can be shown that $c_n(x, d) - c(x, d) \xrightarrow{p} 0$ uniformly in (x, d). Now,

$$c_n(X_j, d_j) - c(X_j, d_j)$$
$$= [\gamma_n(X_j) - \gamma(X_j)] - R_n'M_n^{-1}[\pi_{nj} - (d_j - f_2(X_j))]$$
$$+ (R'M^{-1} - R_n'M_n^{-1})(d_j - f_2(X_j)). \quad \text{(A.8)}$$

The first term in (A.8) converges in probability to 0 uniformly in x. The third term also converges in probability to 0 uniformly in (x, d), because d_j and $f_2(x)$ are bounded, $R_n \xrightarrow{p} R_n$, and $M_n \xrightarrow{p} M$, and M^{-1} exists by assumption. Turning to the second term in (A.8), since $R_n \xrightarrow{p} R$ and $M_n^{-1} \xrightarrow{p} M^{-1}$, it suffices to show that $\pi_{nj} - (d_j - f_2(X_j)) \xrightarrow{p} 0$ uniformly in (X_j, d_j). Now,

$$\pi_{nj} - (d_j - f_2(X_j))$$
$$= \xi_j - \sum_k W_{jk}\xi_k - (d_j - f_2(X_j))$$
$$= (d_j - f_{2n}(X_j)) - \sum_k W_{jk}(d_k - f_{2n}(X_k)) - (d_j - f_2(X_j)),$$

so

$$\sup_{(X_j,d_j)}|\pi_{nj} - (d_j - f_2(X_j))| \le \sup_x|f_2(x) - f_{2n}(x)|$$
$$+ \sup_{(X_j,d_j)}\left|\sum_k W_{jk}(d_k - f_{2n}(X_k))\right|. \quad \text{(A.9)}$$

The first term in (A.9) vanishes by uniform consistency. Now,

$$\sum_k W_{jk}d_k - f_{2n}(X_j) = \sum_k \left[\frac{w((X_j - X_k)/b_n)d_k}{\sum_k w((X_j - X_k)/b_n)}\right]\left[\frac{h_n(X_j)}{h_n(X_k)} - 1\right]$$
$$\le |f_{2n}(X_j)|\sup_{(x,x')}|h_n(x)/h_n(x') - h(x)/h(x')|$$
$$+ \left|\sum_k W_{kj}d_k[h(X_j)/h(X_k) - 1]\right|,$$

with probability 1. The first term converges to 0 in probability because of the uniform consistency of $h_n(x)$, because $f_2(x)$ is bounded, and because $h(x)$ is bounded above and below. Since $h(x)$ is continuous, the second term also vanishes in probability uniformly in (X_j, d_j), by the uniform consistency of kernel regression. Thus $\sum_k W_{jk}d_k - f_{2n}(X_j) \xrightarrow{p} 0$ uniformly in (X_j, d_j). A parallel argument shows that $\sum_k W_{jk}f_{2n}(X_k) - f_{2n}(X_j) \xrightarrow{p} 0$ uniformly in (X_j, d_j). From these two results, it follow that $\sum_k W_{jk}(d_k -$

$f_{2n}(X_k)) \xrightarrow{p} 0$ uniformly in (X_j, d_j). Thus from (A.9) $\pi_{nj} - (d_j - f_2(X_j)) \xrightarrow{p} 0$ uniformly in (X_j, d_j), so the second term in (A.8) converges to 0 uniformly in (X_j, d_j). It now follows from (A.8) that $\sup_{(x,d)}|c_n(x, d) - c(x, d)| \xrightarrow{p} 0$, so $n^{1/2}[B_n - E(B_n|\{X_i, d_i\})] = n^{-1/2} \sum_j c(X_j, d_j)u_j + o_p(1)$.

The second step entails checking Lyapunov's conditions for the central limit theorem (e.g., White 1984) for the independent random variables $Z_j = c_j u_j$: (a) $\text{var}(Z_j) > 0$ and (b) $\exists\ \delta$ and Δ such that $E|Z_j|^{2+\delta} < \Delta < \infty$ for all j. Condition (a) is satisfied, because

$$\begin{aligned} V &= \text{var}(Z_j) \\ &= E(\sigma^2(X_j, d_j)[\gamma(X_j) - R'M^{-1}(d_j - f_2(X_j))]^2) \\ &\geq \underline{\sigma}^2 E\{E[(\gamma(X_j) - R'M^{-1}(d_j - f_2(X_j)))^2|X_j]\} \\ &= \underline{\sigma}^2[E\gamma(X_j)^2 + R'M^{-1}E(d_j - f_2(X_j))(d_j - f_2(X_j))'M^{-1}R] \\ &= \underline{\sigma}^2[E\gamma(X_j)^2 + R'M^{-1}R], \qquad \text{(A.10)} \end{aligned}$$

where the final two equalities have been obtained using $E(d_j - f_2(X_j) \mid X_j) = 0$ and the definition of M. Since M is positive definite by assumption, $\underline{\sigma}^2 > 0$, and $E\gamma(X_j)^2 > 0$ by Assumption 1, it follows that $\text{var}(Z_j) \geq \underline{\sigma}^2\ E\gamma(X_j)^2 > 0$ for all j. Condition (b) follows from Assumption 1(a), the boundedness of $c(x, d)$, and the assumption that u_j has $2 + \delta$ moments that are uniformly bounded in j. Using the equalities in (A.10), if $\sigma^2(x, d) = \sigma^2$ for all (x, d), then

$$\begin{aligned} V &= \sigma^2[E\gamma(X_j)^2 + R'M^{-1}R] \\ &= \sigma^2\left[\int (h^*(x)/h(x) - 1)^2 dH(x) + R'M^{-1}R\right]. \end{aligned}$$

Finally, because $c_n(x, d) - c(x, d) \xrightarrow{p} 0$ uniformly in (x, d) and $E|u_j|^{2+\delta} \leq \Delta < \infty$, $V_n = n^{-1} \sum_j c_n(X_j, d_j)^2 u_{nj}^2 \xrightarrow{p} V$.

[Received December 1985. Revised November 1988.]

REFERENCES

Bierens, H. J. (1983), "Uniform Consistency of Kernel Estimators of a Regression Function Under Generalized Conditions," *Journal of the American Statistical Association*, 77, 699–707.

——— (1985), "Kernel Estimators of Regression Functions," Research Memorandum 8518, University of Amsterdam, Dept. of Economics.

Devroye, L. (1978), "The Uniform Convergence of the Nadaraya–Watson Function Estimate," *Canadian Journal of Statistics*, 6, 179–191.

Devroye, L., and Penrod, C. S. (1984), "The Consistency of Automatic Kernel Density Estimates," *The Annals of Statistics*, 12, 1231–1249.

Devroye, L., and Wagner, T. J. (1980), "Distribution-Free Consistency Results in Non-Parametric Discrimination and Regression Function Estimation," *The Annals of Statistics*, 8, 231–239.

Elbadawi, I., Gallant, A. R., and Souza, G. (1983), "An Elasticity Can be Estimated Consistently Without A Priori Knowledge of Functional Form," *Econometrica*, 51, 1699–1730.

Gallant, A. R. (1981), "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form," *Journal of Econometrics*, 15, 211–245.

Hall, P., and Marron, J. S. (1988), "Choice of Kernel Order in Density Estimation," *The Annals of Statistics*, 16, 161–173.

Harrison, D., Jr., and Rubinfeld, D. (1978), "Hedonic Housing Prices and the Demand for Clean Air," *Journal of Environmental Economics and Management*, 5, 81–102.

Li, K.-C. (1984), "Consistency For Cross-Validated Nearest Neighbor Estimates in Nonparametric Regression," *The Annals of Statistics*, 12, 230–240.

Marron, J. S. (1985), "An Asymptotically Efficient Solution to the Bandwidth Problem of Kernel Density Estimation," *The Annals of Statistics*, 13, 1011–1023.

Nadaraya, E. A. (1964), "On Estimating Regression," *Theory of Probability and Its Applications*, 9, 141–142.

Polinsky, A. M., and Shavell, S. (1975), "The Air Pollution and Property Value Debate," *Review of Economics and Statistics*, 57, 100–104.

——— (1976), "Amenities and Property Values in a Model of an Urban Area," *Journal of Public Economics*, 5, 119–129.

——— (1978), "Amenities and Property Values in a Model of an Urban Area: A Reply," *Journal of Public Economics*, 9, 111–112.

Powell, J. L., Stock, J. H., and Stoker, T. M. (in press), "Semiparametric Estimation of Weighted Average Derivatives," *Econometrica*, 57.

Prakasa Rao, B. L. S. (1983), *Nonparametric Functional Estimation*, New York: Academic Press.

Rice, J. (1984a), "Bandwidth Choice for Nonparametric Regression," *The Annals of Statistics*, 12, 1215–1230.

——— (1984b), "Boundary Modification for Kernel Regression," *Communications in Statistics, Part A—Theory and Methods*, 13, 893–900.

Robinson, P. M. (1988), "Root-*N* Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

Scotchmer, S. (1985), "The Short-Run and Long-Run Benefits of Environmental Improvement," Discussion Paper 1135, Harvard Institute of Economic Research.

Spiegelman, C., and Sacks, J. (1980), "Consistent Window Estimation in Nonparametric Regression," *The Annals of Statistics*, 8, 240–246.

Stock, J. H. (in press), "Nonparametric Policy Analysis: An Application to Estimating Hazardous Waste Cleanup Benefits," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, eds. W. Barnett, J. Powell, and G. Tauchen, Cambridge, U.K.: Cambridge University Press.

Stone, C. J. (1977), "Consistent Nonparametric Regression" (with discussion), *The Annals of Statistics*, 5, 595–645.

Wahba, G. (1978), "Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression," *Journal of the Royal Statistical Society*, Ser. B, 40, 364–372.

Watson, G. S. (1964), "Smooth Regression Analysis," *Sankhyā*, Ser. A, 26, 359–372.

White, H. (1984), *Asymptotic Theory for Econometricians*, Orlando, FL: Academic Press.