*The Annals of Statistics*
1983, Vol. 11, No. 4, 1136–1141

# ON THE CONSISTENCY OF CROSS-VALIDATION IN KERNEL NONPARAMETRIC REGRESSION[1]

BY WING HUNG WONG

*University of Chicago*

For the nonparametric regression model $Y(t_i) = \theta(t_i) + \varepsilon(t_i)$ where $\theta$ is a smooth function to be estimated, $t_i$'s are nonrandom, $\varepsilon(t_i)$'s are i.i.d. errors, this paper studies the behavior of the kernel regression estimate

$$\hat{\theta}(t) = \left[\sum_{j=1}^{n} K\left(\frac{t_j - t}{\lambda}\right) Y(t_j)\right] \Big/ \left[\sum_{j=1}^{n} K\left(\frac{t_j - t}{\lambda}\right)\right]$$

when $\lambda$ is chosen by cross-validation on the average squared error. Strong consistency in terms of the average squared error is established for uniform spacing, compact kernel and finite fourth error moment.

**1. Introduction.** Cross-validation as a method for choosing between estimators or predictors was formally formulated in the articles of Stone (1974) and Geisser (1975). Wahba and Wold (1975) independently proposed cross-validation for choosing the degree of smoothing in spline nonparametric regression. Craven and Wahba (1979), Speckman (1982) discuss consistency of the cross-validated smoothing spline, and Chow, Geman and Wu (1982) establish consistency of kernel density estimator with cross-validation on the likelihood. The present paper investigates consistency of cross-validation for the kernel regression method of Nadaraya (1964) and Watson (1964). Under some regularity conditions, it is proved that the method is consistent in the sense of average square error. There is an extensive literature on the asymptotic properties of kernel regression estimators with *deterministic* sequence of "bandwidth"; see Stone (1979), Devroye and Wagner (1980), Mack and Silverman (1982).

Let observations on the "unknown regression function" $\theta(\cdot)$ be denoted by

$$Y(t_i) = \theta(t_i) + \varepsilon(t_i) \tag{1}$$

where $\varepsilon(t_i)$ are i.i.d. with $E(\varepsilon_i) = 0$ and $\operatorname{Var}(\varepsilon_i) = \sigma^2$. With no model for $\theta(\cdot)$, the nonparametric kernel estimate is defined by

$$\hat{\theta}(t) = G(t)^{-1} \sum_{j=1}^{n} K[(t_j - t)/\lambda] Y(t_j), \tag{2}$$

where $K(\cdot)$ is a symmetric unimodal function, and $G(t) = \sum_{j=1}^{n} K[(t_j - t)/\lambda]$. This estimator was proposed independently by Nadaraya (1964) and Watson (1964). The function $K$ is the kernel function, the parameter $\lambda$ is the "bandwidth" which controls the degree of averaging. Freedman (1981) has emphasized the distinction between the regression model (specified in (1)) and the correlation model under which $\{(Y_i, t_i), i = 1, \cdots, n\}$ is regarded as a random sample from a bivariate distribution. For the correlation model, Efron (1982) gives an excellent discussion of the structure of the prediction problem, cross-validatory assessment, as well as novel improvements upon the cross-validation.

The cross-validatory choice of $\lambda$, based on the average squared prediction error, is that value $\lambda_n^*$ which minimizes

$$CV(\lambda) = n^{-1} \sum_{i=1}^{n} [\hat{\theta}_{-i}(t_i) - Y(t_i)]^2 \tag{3}$$

Received August 1982; revised March 1983.

[1] Support for this research was provided by National Science Foundation Grant No. MCS-8101836.

*AMS* 1980 *subject classification.* Primary, 62G05.

*Key words and phrases.* Cross-validation, consistency, nonparametric regression, kernel estimate.

where $\hat{\theta}_{-i}(t)$ is the estimate of $\theta(t_i)$ computed from the data omitting $Y(t_i)$, i.e.

$$\hat{\theta}_{-i}(t) = G_i(t)^{-1} \sum_{j \neq i} K[(t_j - t)/\lambda] Y(t_j), \quad G_i(t) = \sum_{j \neq i} K[(t_j - t)/\lambda]. \tag{4}$$

The cross-validation function $\mathrm{CV}(\lambda)$ measures the average ability of $\hat{\theta}_{-i}(t_i)$ to predict the "new" observation $Y(t_i)$. Note that forcing $\hat{\theta}$ to do well with respect to prediction error is equivalent to forcing it to estimate $\theta$ well.

**2. Statement of results.** The consistency of the cross-validated kernel estimator with respect to the average squared error loss $n^{-1} \sum_{i=1}^{n} [\hat{\theta}(t_i) - \theta(t_i)]^2$ will be established under the following conditions: (A1): $t_i = i/n$, $i = 1, \cdots, n$; (A2): $K$ is positive, symmetric, with maximum at 0, and is Lipschitz continuous on its support in $[-1, 1]$; (A3): $E(\varepsilon(t_i)^4) < \infty$; (A4): $\theta$ is Hölder continuous, with exponent in $[0, 1]$. The main ideas in the proof can be outlined as follows. If $\lambda^*$ is the CV choice of $\lambda$, and if $\mathrm{ERR}(\lambda) = n^{-1} \sum_{i=1}^{n} [\hat{\theta}(t_i, \lambda) - \theta(t_i)]^2$, then we shall prove $\mathrm{ERR}(\lambda_n^*) \to 0$. Define $\widetilde{\mathrm{ERR}}(\lambda) = n^{-1} \sum_{i=1}^{n} [\hat{\theta}_{-i}(t_i, \lambda) - \theta(t_i)]^2$, which is related to the cross-validation function (3) by

$$\mathrm{CV}(\lambda) = \widetilde{\mathrm{ERR}}(\lambda) + n^{-1} \sum_{i=1}^{n} \varepsilon(t_i)^2 - 2n^{-1} \sum_{i=1}^{n} [\hat{\theta}_{-i}(t_i, \lambda) - \theta(t_i)]\varepsilon(t_i). \tag{5}$$

By the definition of $\lambda^*$, for any deterministic sequence $\lambda_n$,

$$\mathrm{CV}(\lambda_n^*) \le \mathrm{CV}(\lambda_n). \tag{6}$$

Now suppose,

LEMMA 1. *There exists a deterministic sequence $\lambda_n$ such that* $\widetilde{\mathrm{ERR}}(\lambda_n) \to 0$ a.s.

LEMMA 2. *Let* $g(n, \lambda) = n^{-1} \sum_{i=1}^{n} [\hat{\theta}_{-i}(t_i, \lambda) - \theta(t_i)]\varepsilon(t_i)$, then $\sup_{\lambda \in (0,\infty)} |g(n, \lambda)| \to 0$ a.s.

LEMMA 3. $n\lambda_n^* \to \infty$ a.s.

The proofs of these lemmas will be given later. By (5), (6) and Lemmas 1 and 2, it follows that

$$\widetilde{\mathrm{ERR}}(\lambda_n^*) \to 0 \text{ a.s.} \tag{7}$$

The next idea is that $\widetilde{\mathrm{ERR}}$ approximates ERR, so that (7) leads to the following consistency result.

THEOREM. $\mathrm{ERR}(\lambda_n^*) \equiv n^{-1} \sum_{i=1}^{n} [\hat{\theta}(t_i, \lambda_n^*) - \theta(t_i)]^2 \to 0$ a.s.

PROOF. From (2), (4), $\hat{\theta}_{-i}(t_i) - \hat{\theta}(t_i) = K(0)G(t_i)^{-1} [\hat{\theta}_{-i}(t_i) - Y(t_i)]$. Now $G(t_i) = \sum_{|j-i| \le n\lambda} K[(t_j - t_i)/\lambda] \ge \sum_{0 \le \alpha \le n\lambda} K[\alpha/(n\lambda)]$. Since the last term is of order $n\lambda_n^*$, Lemma 3 implies that $G(t_i) \to \infty$ uniformly in $i$. Hence

$$n^{-1}\Sigma[\hat{\theta}_{-i}(t_i, \lambda^*) - \hat{\theta}(t_i, \lambda^*)]^2 \le [\min_{1 \le i \le n} G(t_i)^2]^{-1} K(0)^2 \mathrm{CV}(\lambda_n^*) \to 0 \text{ a.s.}$$

This implies that $\mathrm{ERR}(\lambda_n^*)$ is indeed well approximated by $\widetilde{\mathrm{ERR}}(\lambda_n^*)$, which together with (7) yields the desired result.

It remains to prove the lemmas.

**3. Proof of the lemmas.**

PROOF OF LEMMA 1. Write $\widetilde{\mathrm{ERR}}(\lambda_n)$ as

$$\widetilde{\mathrm{ERR}}(\lambda_n) = n^{-1} \sum_{i=1}^{n} \{\hat{\theta}_{-i}(t_i; \lambda_n) - [G(t_i)/G_i(t_i)]\hat{\theta}(t_i; \lambda_n) + [G(t_i)/G_i(t_i)][\hat{\theta}(t_i; \lambda_n) - \theta(t_i)] + \theta(t_i)[G(t_i)/G_i(t_i) - 1]\}^2,$$

then

$$\widetilde{\mathrm{ERR}}(\lambda_n) \le 4n^{-1} \sum_{i=1}^{n} [K(0)Y(t_i)/G_i(t_i)]^2$$
$$+ 4[\sup_i G^2(t_i)/G_i^2(t_i)]n^{-1} \sum_i [\hat{\theta}(t_i; \lambda_n) - \theta(t_i)]^2 \tag{8}$$
$$+ 4[\sup_i \theta(t_i)]n^{-1} \sum_i [G(t_i)/G_i(t_i) - 1]^2.$$

To treat the first series on the right, note that

$$\limsup_{n\to\infty} n^{-1} \sum_{i=1}^{n} Y^2(t_i) < \infty \quad \text{a.s. under (A3),} \quad \text{while}$$
$$\sup_i[1/G_i(t_i)] \le 1/\inf_i \sum_{j=1}^{n} K[(j-i)/n\lambda_n] \to 0$$

if $\lambda_n \to 0$ and $n\lambda_n \to \infty$. To handle the last term on the right in (8), observe that

$$n^{-1} \sum_i [G(t_i) - G_i(t_i)]^2/G_i^2(t_i) \le K^2(0)\sup_i[1/G_i^2(t_i)] \to 0.$$

To prove that the middle term on the right of (8) converges to zero, observe that

$$\hat{\theta}(t_i; \lambda_n) - \theta(t_i) = A_i + B_i$$

with

$$A_i = G(t_i)^{-1} \sum_j K((t_j - t_i)/\lambda_n)\theta(t_j) - \theta(t_i)$$
$$B_i = G(t_i)^{-1} \sum_j K((t_j - t_i)/\lambda_n)\varepsilon(t_j).$$

By continuity of $\theta(\cdot)$, $A_i \to 0$ uniformly in $i$, thus it suffices to show that $n^{-1}S = n^{-1} \sum i\, B_i^2 \to 0$ a.s.

Write

$$S = 2 \sum_{i<j}\sum \varepsilon(t_i)\varepsilon(t_j)a_{ij} + \sum_i \varepsilon^2(t_i)a_{ii}, \quad \text{where}$$
$$a_{ij} = \sum_k K[(t_k - t_i)/\lambda_n]K[(t_k - t_j)/\lambda_n]/G^2(t_k).$$

Then $E(S - ES)^2 = 4 \sum_{i<j}\sum E[\varepsilon^2(t_i)\varepsilon^2(t_j)]a_{ij}^2 + \sum_i E[\varepsilon(t_i)^2 - \sigma^2]^2 a_{ii}^2 \le \text{const.} \sum_i \sum_j a_{ij}^2$.
Now, $0 \le a_{ij} \le K^2(0) \sum_k G^{-2}(t_k)$, whence

$$n^{-2} \sum_i \sum_j a_{ij} \le \text{const.} [\textstyle\sum_k G^{-2}(t_k)]^2 \le \text{const.} [n(n\lambda_n)^{-2}]^2 = \text{const.}\ n^{-2}\lambda_n^{-4}.$$

Therefore if $\lambda_n = n^{-\alpha}$ where $\alpha < 1/4$ then

$$P(n^{-1} | S - ES | > \varepsilon) \le \text{const.}\ \varepsilon^{-2}n^{-2+4\alpha},$$

whence $n^{-1} | S - ES | \to 0$ a.s., and so $n^{-1}S \to 0$ a.s. Combining the estimates above we see that $\widetilde{\mathrm{ERR}}(\lambda_n) \to 0$, provided $\lambda_n = n^{-\alpha}$ with $\alpha < 1/4$.

PROOF OF LEMMA 2. The proof of Lemma 2 follows from the next two lemmas which involve the following six conditions:

CONDITION (i). Let $\bar{Y}_i = (n-1)^{-1} \sum_{j\ne i} Y_j$, $g(n, \infty) = n^{-1} \sum_i [\bar{Y}_i - \theta(t_i)]\varepsilon(t_i)$, then $g(n, \lambda) \to g(n, \infty)$ uniformly in $n$, as $\lambda \to \infty$.

CONDITION (ii). $g(n, \infty) \to 0$ as $n \to \infty$.

CONDITION (iii). Let $G_\lambda = \sum_{0<|j|\le n\lambda} K(j/n\lambda)$, $\tilde{\theta}_{-i}(t_i, \lambda) = G_\lambda^{-1} \sum_{j\ne i} K[(t_j - t_i)/\lambda]Y(t_j)$, and $\tilde{g}(n, \lambda) = n^{-1} \sum_i [\tilde{\theta}_{-i}(t_i; \lambda) - \theta(t_i)]\varepsilon(t_i)$, then

$$| \tilde{g}(n, \lambda) - g(n, \lambda) | \to 0 \text{ uniformly in } n, \text{ as } \lambda \to 0.$$

CONDITION (iv). $\tilde{g}(n, \lambda) \to 0$ uniformly in $\lambda \le 1$, as $n \to \infty$.

CONDITION (v). $g(n, \lambda) \to 0$ as $n \to \infty$, for any rational $\lambda > 0$.

CONDITION (vi). For any given $\gamma_1 > \gamma_0 > 0$, $\{g(n, \lambda), n = 1, 2, \cdots\}$ is an equicontinuous family of functions of $\lambda \in [\gamma_0, \gamma_1]$.

LEMMA 2A. *For any realization* $\omega$, *conditions* (i) *to* (vi) *imply that* $\sup_{\lambda\in(0,\infty)} |g(n, \lambda)| \to 0$.

PROOF. Let $\varepsilon > 0$ be arbitrary. By (i) and (ii), there exist $n_1$, $\gamma_1$ such that $n > n_1$, $\lambda > \gamma_1 \Rightarrow |g(n, \lambda)|\ \varepsilon$.

By (iii) and (iv), there exist $n_2$, $\gamma_0$ such that $n > n_2$, $\lambda < \gamma_0 \Rightarrow |g(n, \lambda)| \le \varepsilon$.

By (v) and (vi), there exists $n_3$ such that $n > n_3 \Rightarrow \sup_{\lambda\in[\gamma_0,\gamma_1]} |g(n, \lambda)| \le \varepsilon$.

LEMMA 2B. *With probability* 1, *conditions* (i) *to* (vi) *hold simultaneously.*

PROOF. First of all, it is easily verified that $n^{-1}\Sigma |Y_i|$, $n^{-1}\Sigma\varepsilon(t_i)^2$, etc. converges a.s. For (i) it is enough to check that $\limsup n^{-1}\Sigma |Y_i| < \infty \Rightarrow |\hat{\theta}_{-1}(t_i, \lambda) - \bar{Y}_i| \to 0$ uniformly in $n$ and $i$ as $\lambda \to \infty$. For (ii), $g(n, \infty) = n^{-1}\sum_i [(n-1)^{-1} \sum_{j\ne i} (\theta(t_j)) - \theta(t_i)]\varepsilon(t_i) + n^{-1} \Sigma_i [(n-1)^{-1} \sum_{j\ne i} \varepsilon(t_j)]\varepsilon(t_i)$, and both terms can be shown to converge to zero by calculating moments and applying Chebyshev's inequality, as done in the proof of Lemma 1. For (iii), note that $\tilde{\theta}_{-i} = \hat{\theta}_{-i}$ if $n\lambda < i < n - n\lambda$, so that

$$|\tilde{g}(n, \lambda) - g(n, \lambda)| \le [n^{-1} \textstyle\sum_{i\le n\lambda, \text{ or } i\ge n-n\lambda} (\tilde{\theta}_{-i} - \hat{\theta}_{-i})^2]^{1/2}[n^{-1} \sum_{i\le n\lambda, \text{ or } i\ge n-n\lambda} \varepsilon(t_i)^2]^{1/2}.$$

With probability 1, the first factor is uniformly (in $n$) bounded, and the second factor is uniformly (in $n$) convergent to zero as $\lambda \to 0$, since $n^{-1}\Sigma\varepsilon(t_i)^2 \to \sigma^2$ a.s.

For (iv),

$$\begin{aligned}\tilde{g}(n, \lambda) &= n^{-1}\textstyle\sum_i [G_\lambda^{-1} \sum_{j\ne i} K((j-i)/n\lambda)\theta(t_j) - \theta(t_i)]\varepsilon(t_i)\\ &\quad + n^{-1} \textstyle\sum_i [G_\lambda^{-1} \sum_{0<|j-i|\le n\lambda} K((j-i)/n\lambda)\varepsilon(t_j)]\varepsilon(t_i).\end{aligned}$$

Under the assumed continuity of $\theta$, the first term clearly $\to 0$ a.s. With

$$Z_\alpha = n^{-1} \textstyle\sum_{1\le i, i+\alpha\le n} \varepsilon(t_{i+\alpha})\varepsilon(t_i),$$

the second term in the above expression for $\tilde{g}(n, \lambda)$ can be rewritten as $G_\lambda^{-1} \sum_{0<|\alpha|\le n\lambda} K(\alpha/n\lambda)Z_\alpha$, which can again be shown to converge to zero by calculating moments and using Chebyshev's inequality.

For (v), that this holds a.s. is trivial to verify.

For (vi), write

$$\begin{aligned}|\hat{\theta}_{-i}(t_i, \lambda_1) - \hat{\theta}_{-i}(t_i, \lambda_2)| &\le \textstyle\sum_{j\ne i} |Y_j| B_{ij} \quad \text{where}\\ B_{ij} &= |G_i(\lambda_1)^{-1}K[(t_j - t_i)/\lambda_1] - G_i(\lambda_2)^{-1}K[(t_j - t_i)/\lambda_2]|\\ &\le [G_i(\lambda_1)G_i(\lambda_2)]^{-1} \textstyle\sum_{l\ne i} |K[(t_j - t_i)/\lambda_1]K[(t_l - t_i)/\lambda_2]\\ &\quad - K[(t_j - t_i)/\lambda_2]K[(t_l - t_i)/\lambda_1]|\\ &\le 2[G_i(\lambda_1)G_i(\lambda_2)]^{-1} (n-1)\|K\|_C \mathrm{Lip}(K)|\lambda_1^{-1} - \lambda_2^{-1}|\end{aligned}$$

with $\|K\|_C$ and $\mathrm{Lip}(K)$ the sup-norm and the Lipschitz constant of $K$ respectively. Hence, for $\lambda_1, \lambda_2 \ge \gamma_0$, there is a constant $C > 0$ such that $|B_{ij}| \le (C/n\gamma_0^2)|\lambda_1 - \lambda_2|$, $\forall\ i, j$. Hence

$$|\hat{\theta}_{-i}(t_i, \lambda_1) - \hat{\theta}_{-1}(t_i, \lambda_2)| \le C\gamma_0^{-2}|\lambda_1 - \lambda_2|(n^{-1}\Sigma |Y_j|) \to 0$$

uniformly for $n$, $i$ and $\lambda_1 \geq \gamma_0$, as $\lambda_2 \to \lambda_1$ (provided limsup $n^{-1}\Sigma\,|\,Y_j\,| < \infty$, which is true a.s.). Finally,

$$|\,g(n, \lambda_1) - g(n, \lambda_2)\,| \leq n^{-1} \sum_i |\,\hat{\theta}_{-i}(t_i, \lambda_1) - \hat{\theta}_{-i}(t_i, \lambda_2)\,|\,\varepsilon_i \to 0$$

uniformly for $\lambda_1 \geq \gamma_0$, as $\lambda_2 \to \lambda_1$. This completes the proof.

PROOF OF LEMMA 3. We will show that, with probability 1, the supposition that $\tilde{n}\lambda^*_{\tilde{n}} \to \mathrm{M} < \infty$ for any subsequences $\tilde{n}$ will contradict the definition of $\lambda^*$; so suppose $\tilde{n}\lambda^*_{\tilde{n}} \to M$. Now $\hat{\theta}_{-i}(t_i, \lambda^*_{\tilde{n}}) - Y(t_i) = A_i + B_i$ with

$$A_i = G_i(t_i)^{-1}[\textstyle\sum_{j\neq i} K((t_j - t_i)/\lambda^*_{\tilde{n}})\theta(t_j)] - \theta(t_i),$$

$$B_i = G_i(t_i)^{-1}[\textstyle\sum_{j\neq i} K((t_j - t_i)/\lambda^*_{\tilde{n}})\varepsilon(t_j)] - \varepsilon(t_i).$$

It is clear that $A_i \to 0$ uniformly in $i$ as $n \to \infty$. By the same argument as in the proof of condition (iii) of Lemma 2, we can replace $G_i$ by $G_\lambda$; hence for $n = \tilde{n}$ and $\lambda = \lambda^*_{\tilde{n}}$,

$$\begin{aligned} n^{-1} \textstyle\sum_i B_i^2 &= n^{-1} \textstyle\sum_i \varepsilon(t_i)^2 - 2n^{-1} \textstyle\sum_i \varepsilon(t_i)[G_\lambda^{-1} \textstyle\sum_{|j-i|\leq n\lambda, j\neq i} K((j-i)/n\lambda)\varepsilon(t_j)] \\ &\quad + n^{-1} \textstyle\sum_i [\textstyle\sum_{|j-i|\leq n\lambda, j\neq i} \textstyle\sum_{|j'-i|\leq n\lambda, j'\neq i} G_\lambda^{-2} K((j-i)/n\lambda) \\ &\qquad \cdot K((j'-i)/n\lambda)\varepsilon(t_j)\varepsilon(t_{j'})] \\ &= n^{-1} \textstyle\sum_i \varepsilon(t_i)^2 - 2 \textstyle\sum_{0<|\alpha|\leq n\lambda} G_\lambda^{-1} K(\alpha/n\lambda)[n^{-1} \textstyle\sum_{1\leq i+\alpha, i\leq n} \varepsilon(t_{i+\alpha})\varepsilon(t_i)] \\ &\quad + \textstyle\sum_{0<|\alpha|\leq n\lambda} G_\lambda^{-2} K^2(\alpha/n\lambda)[n^{-1} \textstyle\sum_{1\leq i+\alpha, i\leq n} \varepsilon(t_{i+\alpha})^2] \\ &\quad + \textstyle\sum_{0<|\alpha|\leq n\lambda} \textstyle\sum_{\substack{0<|\alpha'|\leq n\lambda \\ \alpha\neq\alpha'}} G_\lambda^{-2} K(\alpha/n\lambda)K(\alpha'/n\lambda) \\ &\qquad \cdot [n^{-1} \textstyle\sum_{1\leq i+\alpha\leq n, 1\leq i+\alpha'\leq n} \varepsilon(t_{i+\alpha'})\varepsilon(t_{i+\alpha})]. \end{aligned}$$

It is clear that $n^{-1} \sum_{1\leq i+\alpha\leq n, 1\leq i+\alpha'\leq n} \varepsilon(t_{i+\alpha})\varepsilon(t_{i+\alpha'}) \to 0$ uniformly for $\alpha$, $\alpha'$ such that $-[M+1] \leq \alpha \neq \alpha' \leq [M+1]$.

But for $\tilde{n}$ large enough $\tilde{n}\lambda^*_{\tilde{n}} \leq [M+1]$, so that along the subsequence $\tilde{n}$,

$$\tilde{n}^{-1} \textstyle\sum_i B_i^2 \to \sigma^2 + \sigma^2 \textstyle\sum_{0\leq|\alpha|\leq M} K^2(\alpha/M)[\textstyle\sum_{0<|j|\leq M} K(j/M)]^{-2} > \sigma^2.$$

Hence lim inf $\mathrm{CV}(\lambda^*) =$ lim inf $n^{-1} \sum_i B_i^2 > \sigma^2$. But this contradicts the definition of $\lambda^*$, since by Lemma 1 one can choose a deterministic sequence $\lambda_n$ such that $\mathrm{CV}(\lambda_n) \to \sigma^2$ a.s.

**Acknowledgement.** I am grateful to a referee whose comments and suggestions were found to be very helpful in the revision of this paper, and to Dennis Cox and K. C. Li for pointing out a mistake in the first draft.

## REFERENCES

CHOW, Y. S., GEMAN, S. and WU, L. D.(1983). Consistent cross-validated density estimation. *Ann. Statist.* **11** 25–38.

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions. *Numer. Math.* **31** 377–404.

DEVROYE, L. P. and WAGNER, T. J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Ann. Statist.* **8** 231–239.

EFRON, B. (1983). Estimating the error rate of a prediction rule: improvement on cross-validation. *J. Amer. Statist. Assoc.* **78** 316–331.

FREEDMAN, D. A. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228.

GEISSER, S. (1975). The predictive sample reuse method with applications. *J. Amer. Statist. Assoc.* **70** 320–328.

MACK, Y. P. and SILVERMAN, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. verw. Gebiete* **61** 405–415.

NADARAYA, E. A. (1964). On estimating regression. *Theor. Probab. Appl.* **9** 141–142.

SPECKMAN, P. L. (1982). Spline smoothing and optimal rates of convergence in nonparametric regression models. Unpublished manuscript.

STONE, C. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645.

STONE, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. Royal Statist. Soc. Ser. B.* **36** 111–147.

WAHBA, G. and WOLD, S. (1975). A completely automatic French curve: fitting spline functions by cross-validation. *Communications in Statistics* **4** 1–17.

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A.* **26** 359–372.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637