



**Discussion: Polynomial Splines and their Tensor Products in Extended Linear Modeling**

W. Hardle; J. S. Marron; L. Yang

*Annals of Statistics*, Volume 25, Issue 4 (Aug., 1997), 1443-1450.

Stable URL:

<http://links.jstor.org/sici?sici=0090-5364%28199708%2925%3A4%3C1443%3ADPSATT%3E2.0.CO%3B2-8>

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

*Annals of Statistics* is published by Institute of Mathematical Statistics. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ims.html>.

---

*Annals of Statistics*

©1997 Institute of Mathematical Statistics

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2003 JSTOR

- O'SULLIVAN, F., YANDELL, B. and RAYNOR, W. (1986). Automatic smoothing of regression functions in generalized linear models. *J. Amer. Statist. Assoc.* **81** 96–103.
- PAWITAN, Y. and O'SULLIVAN, F. (1994). Nonparametric spectral density estimation using penalized Whittle likelihood. *J. Amer. Statist. Assoc.* **89** 600–610.
- SHEN, X. and HU, D. (1994). Universal sieve scheme and spline adaptation. Technical report, Dept. Statistics, Ohio State Univ.
- SILVERMAN, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann. Statist.* **10** 795–810.
- WAHBA, G. (1980). Automatic smoothing of the log periodogram. *J. Amer. Statist. Assoc.* **75** 122–132.
- WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.
- WAHBA, G. and LUO, Z. (1995). Smoothing spline ANOVA fits for very large, nearly regular data sets, with application to historical global climate data. Technical Report 953, Dept. Statistics, Univ. Wisconsin.
- WAHBA, G., WANG, Y., GU, C., KLEIN, R. and KLEIN, B. (1995). Smoothing spline ANOVA for exponential families, with application to the Wisconsin epidemiological study of diabetic retinopathy. *Ann. Statist.* **23** 1865–1895.
- WANG, Y. (1995). Grkpack: fitting smoothing spline ANOVA models for exponential families. Technical Report 942, Dept. Statistics, Univ. Wisconsin. (Available on-line at <http://www.sph.umich.edu/~yuedong>.)
- WANG, Y. (1996a). Smoothing spline models with correlated random errors. Technical report, Dept. Biostatistics, Univ. Michigan. (Available on-line at <http://www.sph.umich.edu/~yuedong>.)
- WANG, Y. (1996b). Mixed-effects smoothing spline ANOVA. *J. Roy. Statist. Soc. Ser. B*. To appear. (Available on-line at <http://www.sph.umich.edu/~yuedong>.)
- WANG, Y. and WAHBA, G. (1995). Bootstrap confidence intervals for smoothing splines and their comparison to Bayesian confidence intervals. *J. Statist. Comput. Simulation* **51** 263–279.
- WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester.
- XIANG, D. and WAHBA, G. (1996). A generalized approximate cross validation for smoothing splines with non-Gaussian data. *Statist. Sinica* **6** 675–692.
- ZUCKER, D. M. and KARR, A. F. (1990). Nonparametric survival analysis with time-dependent covariate effects: a penalized partial likelihood approach. *Ann. Statist.* **18** 329–353.

DEPARTMENT OF STATISTICS  
1399 MATH-SCI BUILDING  
PURDUE UNIVERSITY  
WEST LAFAYETTE, INDIANA 47907-1399  
E-MAIL: chong@stat.purdue.edu

## DISCUSSION

W. HÄRDLE,<sup>1</sup> J. S. MARRON<sup>2</sup> AND L. YANG<sup>1</sup>

*Humboldt Universität zu Berlin, University of North Carolina, Chapel Hill  
and Humboldt Universität zu Berlin*

Stone, Hansen, Kooperberg and Truong have written an excellent review of the fine work they have done in making one type of spline modeling useful

<sup>1</sup>Supported by Sonderforschungsbereich 373 “Quantifikation und Simulation Ökonomischer Prozesse” Deutsche Forschungsgemeinschaft.

<sup>2</sup>Supported by NSF Grant DMS-9504414.

in a wide variety of statistical problems. The unifying framework of *extended linear modeling* provides substantial insights about the essential ideas of this type of spline smoothing.

While the presentation is generally excellent, we question the chosen terminology “polynomial splines.” The problem is that the rather popular “smoothing spline,” which is a solution of a regularization problem and already the subject of two monographs [Wahba (1990) and Green and Silverman (1994)], is also a polynomial spline (although of a very different type). In our view, it would be more appropriate to call the type of spline in the present paper by their older names of “B-spline” or “regression spline.”

This discussion consists of questions in four directions: interpretability, theory versus implementation, the effectiveness of knot deletion algorithms and how applicable the present methodology is to some problems in time series and model testing.

**1. Interpretability.** Many statisticians view simplicity and intuitive understanding of “what the smooth is doing to the data” as very important criteria in choosing a smoothing method. In this respect, we suggest moving average–kernel–local polynomial methods as being preferable, and we believe that they will continue to have an enduring attraction to many statisticians for this reason. We note that Kooperberg and Stone (1991) were not immune to this appeal, and used a very simple kernel method to show (quite convincingly) that the spline method at that time was doing a very good job of density estimation (in fact better than “higher tech” kernel methods). From the point of view of simplicity and interpretability, we ask: “if the kernel method is how one really understands what is going on in the data, why should one then construct the spline?”

**2. Theory versus implementation.** The gap between what is called “the nonadaptive procedures that we can treat analytically” and “the adaptive methodologies that we have implemented” is somewhat worrying. We are unsure about the suggestion that this is merely because the knot deletion/addition is not very tractable to mathematical analysis. Instead we wonder: “has nobody been able to show this adaptive method is statistically efficient because it is inefficient?”

An alternate, intuitively appealing approach to knot choice for B/regression splines, based on Bayes’ ideas, has been developed in Smith and Kohn (1994, 1996a, b). See the Ph.D. dissertation by Smith (1996) for an excellent summary, and a compelling case made for the effectiveness of this approach. A direct comparison of this Bayesian approach with that of the present paper would be quite interesting in terms of statistical efficiency, flexibility and also computation time. We note that Smith and Kohn do much more extensive simulation, and we wonder if this is because their methods are faster to compute.

**3. Knot deletion.** In this section, we look carefully at some ideas which give us doubts concerning the issues raised in Section 2. We focus here on the

one dimensional regression setting, which is probably the easiest to understand and interpret, but the issues we raise here likely exist as well in other settings.

In Section 5 of Stone, Hansen, Kooperberg and Truong, the minimal space is the space of constant functions. Here we show that more discussion of this issue is needed. In particular, we show that without this restriction, the knot deletion procedure can give poor performance both asymptotically and with a simulated example. We wonder if this is an anomaly of the minimal space or if this is what lies at the root of the fact that good asymptotic properties have not been demonstrated for knot deletion methods. In our asymptotics we show that in a “high noise case” a crucial term can be improperly eliminated by knot deletion, which leads to an inconsistent estimate. We then show that this effect is not just an artifact of our asymptotic model by considering a reasonable simulated example where this occurred.

For this, consider the simple regression model

$$Y = m(X) + \varepsilon,$$

where  $X$  is assumed to be uniformly distributed on  $[0, 1]$ ,  $\varepsilon$  is independent of  $X$  and normally distributed with mean 0 and variance  $\sigma^2$ , and  $m(\cdot)$  is a function defined on  $[0, 1]$  with piecewise continuous derivative. Using equally spaced knots initially [as in Stone (1985, 1994)], we let the basis functions be

$$B_0(x) \equiv 1, \quad B_1(x) = x, \quad B_j(x) = \left(x - \frac{j-1}{J}\right)_+, \quad j = 2, \dots, J,$$

where  $J = O(n^{1/3})$  [because here  $p = d = 1$  in the notation of Stone (1994), where  $p$  is the degree of smoothness and  $d$  is the highest degree of interaction allowed]. For simplicity, set  $J = 2\lceil n^{1/3}/2 \rceil + 1$ . Given a random sample  $(X_i, Y_i)_{i=1}^n$ , let  $\hat{m}(x) = \sum_{j=0}^J \hat{\beta}_j B_j(x)$  be the linear spline estimator of  $m(\cdot)$ , where the coefficients  $\hat{\beta}_j$  satisfy

$$(\hat{\beta}_j) = \arg \min_{\beta \in \mathbb{R}^{J+1}} \frac{1}{n} \sum_{i=1}^n \left(Y_i - \sum_{j=0}^J \beta_j B_j(X_i)\right)^2.$$

The deletion rule in the regression setting is to use the residual sum of squares to decide which basis function to add or delete. According to the definition of allowable spaces, the function  $B_1$  cannot be deleted unless all  $B_j$ ,  $j \geq 2$ , had been deleted. Although new knots at preselected order statistics of the data could be added, we believe that the addition of these new knots would not significantly change the situation that we are addressing here. We let  $\hat{\beta}_j$ ,  $j = 0, 1, 2, \dots, J$ , denote the estimated coefficients for the allowable space that has the smallest GCV value, after the deletion process is done. An indication of difficulties in this context is given by:

**PROPOSITION 1.** *Under the above assumptions and without the restriction of constant functions being in the minimal space, if  $m(x) = 1+bx$  and  $\sigma = 2\lceil n^{1/3} \rceil$ ,*

there exists a constant  $C_1 > 0$ , such that

$$\liminf_{n \rightarrow \infty} P[\widehat{\beta}_0 = 0] \geq C_1.$$

The fact that this leads to inconsistency is summarized as:

**COROLLARY 1.** *Under the above assumptions and without the restriction of the constant function being in the minimal space, if  $m(x) = 1 + bx$  and  $\sigma = 2[n^{1/3}]$ , there exist constants  $C_1 > 0$  and  $C_2 > 0$ , such that*

$$\liminf_{n \rightarrow \infty} P[\|\widehat{m}(x) - m(x)\|_\infty \geq 1] \geq C_1.$$

The assumption  $\sigma \approx n^{1/3}$  is a model for “high noise with respect to the sample size.” Such noise levels often occur in econometrics. This version of the adaptive spline seems questionable in such applications, because it is inconsistent. This makes us wonder about possible similar inefficiencies in the knot deletion approach because of similar occurrences for other “important” basis functions. Note also that this estimation context is not impossibly difficult. For example, using the simple Nadaraya–Watson estimator, the  $L_\infty$  rate is

$$O[n^{1/3}(n^{-1}h^{-1} \log n)^{1/2} + h],$$

optimized when the bandwidth  $h$  is chosen at the rate  $n^{-1/9}(\log n)^{1/3}$ , which is also the optimal rate. See, for example, Györfi, Härdle, Sarda and Vieu [(1989), Theorem 3.3.0, page 23].

Figure 1 shows a simulated example which demonstrates that the problem of deleting important knots in this way is not an asymptotic oddity. The target function is linear,  $m(x) = 0.6x + 0.2$ . The data come from adding i.i.d.  $N(0, 0.25)$  noise as shown. The estimate comes from doing knot deletion and then finding the AIC and BIC best choices of the number of knots (they were the same for this data set). The poor behavior on the right-hand side is “bad luck” because the data happen to be larger than usual in that area. However, the poor behavior on the left-hand side is caused by the fact that the intercept term of the model is deleted relatively early in the sequence. We are concerned about this because the intercept is actually part of the underlying model.

Our question here is: “are these simply artifacts of our ignoring the minimal space or are they indicators that in fact knot deletion is an inefficient method of adaptation?”

**4. Time series and model testing.** In nonlinear time series analysis, the conditional variance is often of interest, sometimes more than the conditional mean (for some econometrics data, for example). A review of some recent works in this area can be found in Härdle and Chen (1995) and the references therein. Simultaneous estimation of additive mean and multiplicative volatility functions in autoregressive time series has been done with the local polynomial method by Yang and Härdle (1996). Härdle, Tsybakov and Yang (1996) have also developed estimation procedures of the mean and covariance

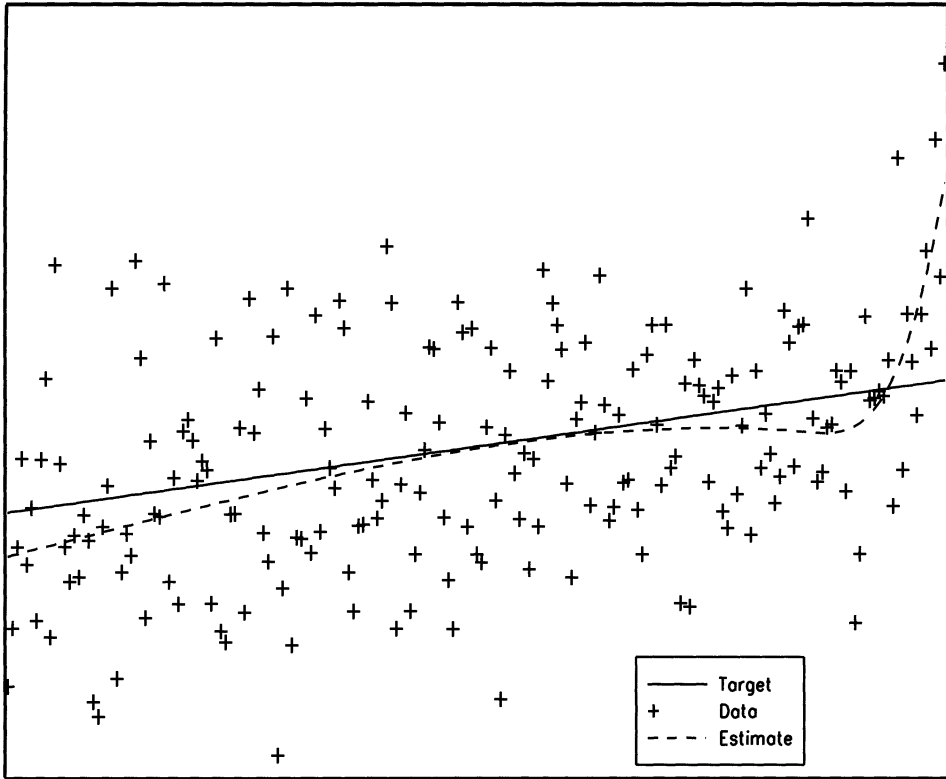


FIG. 1.

function in vector autoregression using local linear estimators. It would be interesting to see such results obtained with the spline approach.

Another area of interesting research where the local polynomial method has been successfully employed is the testing of models. In particular, Härdle, Mammen and Müller (1996) developed procedures for testing parametric versus semiparametric modeling in generalized regression, while Härdle and Yang (1996) developed procedures for testing linearity of main effects in generalized additive regression. Again, it would be interesting to see work in these areas using B/regression splines, which we suspect may be more difficult.

## APPENDIX

**PROOF OF PROPOSITION 1.** We denote by  $\mathbf{X}$  the vector  $(X_1, X_2, \dots, X_n)^T$ , by  $\mathbf{Y}$  the vector  $(Y_1, Y_2, \dots, Y_n)^T$  and by  $\varepsilon$  the vector  $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$ . The inner product on  $\mathbb{R}^n$  is defined as  $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle_n = (1/n) \sum_{i=1}^n x_i y_i$ , while for functions as  $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$ , the norms are defined accordingly. Also denote by  $V_{\{j_1, j_2, \dots, j_k\}}$  the function space spanned by  $B_{j_1}(x), B_{j_2}(x), \dots$ ,

$B_{j_k}(x)$  and by  $\widehat{V}_{\{j_1, j_2, \dots, j_k\}}$  the space spanned by  $B_{j_1}(\mathbf{X}), B_{j_2}(\mathbf{X}), \dots, B_{j_k}(\mathbf{X})$ . For any vector  $v$  (or function  $f$ ), we denote also by  $v_{\{j_1, j_2, \dots, j_k\}}$  (or  $f_{\{j_1, j_2, \dots, j_k\}}$ ) the projection of  $v$  (or  $f$ ) and by  $v_{\{j_1, j_2, \dots, j_k\}}^\perp$  (or  $f_{\{j_1, j_2, \dots, j_k\}}^\perp$ )  $v - v_{\{j_1, j_2, \dots, j_k\}}$  (or  $f - f_{\{j_1, j_2, \dots, j_k\}}$ ). The distance from  $v$  to  $\widehat{V}_{\{j_1, j_2, \dots, j_k\}}$  is  $\|v_{\{j_1, j_2, \dots, j_k\}}^\perp\|$ , which we denote by  $d(v)_{\{j_1, j_2, \dots, j_k\}}$  and so forth. For now, we fix  $m(x) \equiv 1 + bx$ . Without loss of generality, we take  $b = 0$  because the function  $B_1(x) \equiv x$  cannot be deleted.

LEMMA 1. For any  $j = 2, 3, \dots, J$ ,  $1/4 \leq (J/j)D_j \leq 1/3$ , where

$$D_j = d(m(x))_{\{1, j, j+1, \dots, J\}}^2,$$

and consequently

$$\frac{1}{4}(1 + O_p(n^{-1/2})) \leq \frac{J}{j} \mathbf{D}_j \leq \frac{1}{3}(1 + O_p(n^{-1/2}))$$

where  $\mathbf{D}_j = d(m(\mathbf{X}))_{\{1, j, j+1, \dots, J\}}^2$ .

PROOF. It is easy to verify that  $\|m(x) - (J/j)[B_1(x) - B_j(x)]\|^2 = j/(3J)$ , which implies the right-hand side of the inequality. Note that among the functions  $B_1(x), B_j(x), B_{j+1}(x), \dots, B_J(x)$ , only  $B_1(x)$  is nonzero on the interval  $(0, j/J)$ ; thus,

$$d(m(x))_{\{1, j, j+1, \dots, J\}}^2 \geq \min_t \int_0^{j/J} (1 - tx)^2 dx = \frac{j}{4J}. \quad \square$$

LEMMA 2. For any  $j = 2, 3, \dots, J$ ,

$$\|\varepsilon_{\{1, j, j+1, \dots, J\}}\|^2 = \frac{4(J-j)}{n^{1/3}}(1 + o_p(1)).$$

PROOF. The two facts needed for the proof are  $\varepsilon$  is  $N(\mathbf{0}_n, \sigma^2 \mathbf{I}_{n \times n}) = N(\mathbf{0}_n, 4n^{2/3} \mathbf{I}_{n \times n}(1 + o(n^{-1/3}))$  and the projection subspace is of dimension  $1 + (J - j + 1)$ .  $\square$

LEMMA 3. For any  $j = 2, 3, \dots, J$ ,

$$\frac{n^{1/6}}{2\sqrt{\mathbf{D}_j}} \langle \varepsilon, m(\mathbf{X})_{\{1, j, j+1, \dots, J\}}^\perp \rangle \rightarrow N(0, 1).$$

The proof is similar to the previous lemmas.

To complete the proof of Proposition 1, we want to prove that the following event has a probability  $\geq C_1 > 0$ :

$$\begin{aligned} & \|\varepsilon_{\{0, 1, \dots, j-1, j+1, \dots, J\}}\|^2 + \|m(\mathbf{X})_{\{0, 1, \dots, j-1, j+1, \dots, J\}}\|^2 \\ & + 2\langle \varepsilon_{\{0, 1, \dots, j-1, j+1, \dots, J\}}, m(\mathbf{X})_{\{0, 1, \dots, j-1, j+1, \dots, J\}} \rangle \\ & < \|\varepsilon_{\{1, \dots, j-1, j, j+1, \dots, J\}}\|^2 + \|m(\mathbf{X})_{\{1, \dots, j-1, j, j+1, \dots, J\}}\|^2 \\ & + 2\langle \varepsilon_{\{1, \dots, j-1, j, j+1, \dots, J\}}, m(\mathbf{X})_{\{1, \dots, j-1, j, j+1, \dots, J\}} \rangle \end{aligned}$$

for all  $j = 2, 3, \dots, J$ . This is proved by noting that in fact  $m(\mathbf{X})_{\{0, 1, \dots, j-1, j+1, \dots, J\}} = m(\mathbf{X})$ ,

$$\begin{aligned} & \|\varepsilon_{\{0, 1, \dots, j-1, j+1, \dots, J\}}\|^2 - \|\varepsilon_{\{1, \dots, j-1, j, j+1, \dots, J\}}\|^2 \\ & \leq \|\varepsilon_{\{0, 1, \dots, j-1, j, j+1, \dots, J\}}\|^2 - \|\varepsilon_{\{1, \dots, j-1, j, j+1, \dots, J\}}\|^2 \leq \frac{4}{n^{1/3}}(1 + o_p(1)) \end{aligned}$$

by Lemma 1 and

$$\begin{aligned} & \|m(\mathbf{X})_{\{0, 1, \dots, j-1, j+1, \dots, J\}}\|^2 - \|m(\mathbf{X})_{\{1, \dots, j-1, j, j+1, \dots, J\}}\|^2 \\ & = \|m(\mathbf{X})\|^2 - \|m(\mathbf{X})_{\{1, \dots, j-1, j, j+1, \dots, J\}}\|^2 \leq \frac{1}{3J} \end{aligned}$$

by Lemma 2, while

$$\begin{aligned} & 2\langle \varepsilon_{\{0, 1, \dots, j-1, j+1, \dots, J\}}, m(\mathbf{X})_{\{0, 1, \dots, j-1, j+1, \dots, J\}} \rangle \\ & - 2\langle \varepsilon_{\{1, \dots, j-1, j, j+1, \dots, J\}}, m(\mathbf{X})_{\{1, \dots, j-1, j, j+1, \dots, J\}} \rangle \\ & = 2\langle \varepsilon, m(\mathbf{X})_{\{1, \dots, j-1, j, j+1, \dots, J\}}^\perp \rangle \end{aligned}$$

has variance of order  $n^{-1/3}(1/J)$  or  $n^{-2/3}$  by Lemma 3. Therefore the probability of

$$2\langle \varepsilon, m(\mathbf{X})_{\{1, \dots, j-1, j, j+1, \dots, J\}}^\perp \rangle + \frac{4}{n^{1/3}}(1 + o_p(1)) + \frac{1}{3J} < 0$$

is greater than a positive constant, meaning that the constant basis  $B_0$  would be the first one to be removed. It is easy to verify that the GCV for the space  $\widehat{V}_{\{1, 2, \dots, J\}}$  can also be made smaller than that of  $\widehat{V}_{\{0, 1, 2, \dots, J\}}$  with positive probability. In other words, the event ( $\alpha = 2.5$  according to the end of Section 5.2 of Stone, Hansen, Kooperberg and Truong)

$$\begin{aligned} & \left( \frac{n - \alpha J}{n - \alpha J - 1} \right)^2 \|\varepsilon^\perp_{\{0, 1, \dots, J\}}\|^2 \\ & > \|\varepsilon^\perp_{\{1, \dots, J\}}\|^2 + 2\langle \varepsilon^\perp, m(\mathbf{X})_{\{1, \dots, J\}} \rangle + \|m(\mathbf{X})_{\{1, \dots, J\}}^\perp\|^2 \end{aligned}$$

can have a positive probability as well. Thus we have shown that the probability that the final model does not contain the constant term is positive.  $\square$



PROOF OF COROLLARY 1. Note again that among the functions  $B_j(x)$ ,  $1 \leq j \leq J$ , only  $B_1(x)$  is nonzero on the interval  $(0, j/J)$ , thus, when the constant term is not in the final model,

$$\|\widehat{m}(x) - m(x)\|_\infty \geq \inf_{t \in \mathbb{R}^1} \sup_{x \in (0, 1/J)} |1 - tx| = 1. \quad \square$$

## REFERENCES

- GREEN, P. and SILVERMAN, B. W. (1994). *Nonparametric Regression and Generalized Linear Models, a Roughness Penalty Approach*. Chapman and Hall, London.
- GYÖRFI, L., HÄRDLE, W., SARDA, P. and VIEU, P. (1989). *Nonparametric Curve Estimation from Time Series. Lecture Notes in Statist.* **60**. Springer, Berlin.
- HÄRDLE, W. and CHEN, R. (1995). Nonparametric time series analysis, a selective review with examples. In *Proceedings of the 50th Session of the International Statistical Institute, Beijing*. International Statistical Institute, Voorburg.
- HÄRDLE, W., MAMMEN, E. and MÜLLER, M. (1996). Testing parametric versus semiparametric modelling in generalized linear models. Discussion Paper 28, Sonderforschungsbereich 373.
- HÄRDLE, W., TSYBAKOV, A. and YANG, L. (1997). Nonparametric vector autoregression. *J. Statist. Plann. Inf.* To appear.
- HÄRDLE, W. and YANG, L. (1996). Generalized additive models: derivative estimation and hypotheses testing. Discussion paper, Sonderforschungsbereich 373.
- SMITH, M. S. (1996). Nonparametric regression: a Markov chain Monte Carlo approach. Ph.D. dissertation, Univ. New South Wales.
- SMITH, M. and KOHN, R. (1994). A Bayesian approach to additive nonparametric regression. In *Computationally Intensive Statistical Methods: Proceedings of the 26th Symposium on the Interface* (J. Sall and A. Lehman, eds.) **26** 96–104. Interface Foundation, Fairfax Station, VA.
- SMITH, M. and KOHN, R. (1996). Nonparametric regression via Bayesian variable selection. *J. Econometrics* **75** 317–327.
- SMITH, M. and KOHN, R. (1996b). Nonparametric bivariate regression. Unpublished manuscript.
- STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705.
- STONE, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation (with discussion). *Ann. Statist.* **22** 118–184.
- TRUONG, Y. and STONE, C. J. (1992). Nonparametric function estimation involving time series. *Ann. Statist.* **20** 77–97.
- TRUONG, Y. and STONE, C. J. (1994). Semiparametric time series regression. *J. Time Series Anal.* **15** 405–428.
- WAHBA, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.
- YANG, L. and HÄRDLE, W. (1997). Nonparametric autoregression with additive mean and multiplicative volatility. *J. Time Series Anal.* To appear.

W. HÄRDLE  
L. YANG  
INSTITUT FÜR STATISTIK UND ÖKONOMETRIE  
WIRTSCHAFTSWISSENSCHAFTLICHE FAKULTÄT  
HUMBOLDT UNIVERSITÄT ZU BERLIN  
SPANDAUER STRASSE 1  
D-10178 BERLIN  
GERMANY

J. S. MARRON  
DEPARTMENT OF STATISTICS  
UNIVERSITY OF NORTH CAROLINA  
CHAPEL HILL, NORTH CAROLINA 27599