Sperlich, Stefan

Working Paper

About sense and nonsense of non- and semiparametric analysis in applied econometrics

SFB 373 Discussion Paper, No. 2003,36

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Sperlich, Stefan (2003) : About sense and nonsense of non- and semiparametric analysis in applied econometrics, SFB 373 Discussion Paper, No. 2003,36, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,
https://nbn-resolving.de/urn:nbn:de:kobv:11-10050478

This Version is available at:
https://hdl.handle.net/10419/22251

About Sense and Nonsense of Non- and Semiparametric Analysis in Applied Econometrics[1]

Stefan Sperlich

Universidad Carlos III de Madrid, C/Madrid 126, 28903 Getafe

Summary

The discussion about the use of semiparametric analysis in empirical research in economics is as old as the methods are. This article can certainly not be more than a small contribution to the polemic question how useful is non- or semiparametric statistics for applied econometrics. The goal is twofold: to illustrate that the use of these methods have their justification in economics, and to highlight what might be reasons for the lack of its application in empirical research. We do not give a survey of available methods and procedures. Since we discuss the question of the use of non- or semiparametric methods (in economics) in general, we believe that it is fair enough to stick to kernel smoothing methods. It might be that we will face some deficiencies that are more typical in the context of kernel smoothing than it is for other methods. However, the different smoothing methods share mainly the same advantages and disadvantages we will discuss. Even though many points of this discussion hold also true for other research fields, all our examples are either based on economic data sets or concentrate on models that are typically motivated from economic or econometric theory.

[1]This research was supported by the "Dirección General de Enseñanza Superior" BEC2001-1270. We thank J.Mora, L.Collado, and J. Rodríguez-Poó for helpful discussion.

1 Introduction

When a group of researchers specialized in non- and semiparametric statistics, and working since many years mainly in this filed, meet to a workshop "The art of semiparametrics" (moreover, with an explicit section about econometrics), then this seems to be the right forum for a discussion about the following questions: What are the reasons for the continuing lack of applications of non- and semiparametric methods in the empirical research in economics and applied econometrics? Actually, it is not only the lack of application that should concern; often one can even find a strong rejection of these methods from a significant part of the researchers in economics. It is clear from the beginning that it will be impossible to convince those who insist that "these recent developments in statistics are of no use for a better understanding of economic processes" or that there is no need because "all functional forms found by nonparametric methods could have easily been modeled with more conventional parametric ones". Sometimes it is just insufficient mathematical knowledge that causes the dislikes, when e.g. non- and semiparametric methods are considered as "too technical" or when people justify their dislike with the bias inherent in nonparametrics, the lack of knowledge about the degrees of freedom, etc.. In contrast to those "arguments", there are many good reasons why in empirical research, especially in economics, non- or semiparametric applications are rare and many empirical researcher suspicious of these methods.

In several joint works and discussions with different economists, there usually came up the following criticisms:

- lack of interpretability of the estimates, e.g. causalities remain unclear and the lack of possibilities of modeling
- problems with the choice of smoothing parameters (in future SP), and lack of its interpretability
- economic data sets are usually high dimensional and contain many discrete variables; so they have a structure that is hard to manage for nonparametric methods
- imposing restrictions like monotonicity is rather cumbersome
- the treatment of endogeneity and simultaneous equation systems is rather crucial in economics but neglected in the statistic literature
- often, neither the optimal fit nor the regression function on its own are the target of interest
- lack of automatization; the methods are too complex to be managed by the empirical researcher without support from a specialist in nonparametrics

Further, let us recall what Stone (1985) said about the task of statistical modeling. He states that the three fundamental aspects of statistical models are flexibility, dimensionality and interpretability. "Flexibility is the ability of the model to provide accurate fits in a wide variety of situations, inaccuracy here leading to bias in estimation. Dimensionality can be thought of in terms of the variance in estimation, the curse of dimensionality being that the amount of data required to avoid an unacceptable large variance increases rapidly with increasing dimensionality. In practice there is an inevitable trade-off between flexibility and dimensionality or, as usually put, between bias and variance. Interpretability lies in the potential for shedding light on the underlying structure".

Comparing these criteria with the list of criticism from above, we see very nicely how they are interconnected and related to each other. Moreover, we can say: flexibility is given by the nature of nonparametrics; with respect to dimensionality much has been done in the last ten years, but interpretability and "automatization" (including the SP selection) seem to remain the main obstacles.

Obviously, we neither can comment in detail on all these criticisms nor offer solutions to them here. It is evident that most of them refer mainly to the problem of estimation. Indeed, the use of testing methods is much less polemic, except the discussion about optimality and efficiency among statisticians and econometricians.

Due to all this we have decided for the following organization of this paper: We concentrate on the perspectives of the existing non- and semiparametric methods for (research in) economics, discussing briefly some of the open problems where existent. Further, we will separate the discussion of testing from the one of estimation, giving the main emphasis on the second part. The numerous examples provided form the largest part of this article. They are certainly not closed empirical research projects but shall help for illustration. We always try to consider relatively simple regression models (even in the testing part) to highlight our points. So we exclude e.g. transfomration models, measurement error models, survival functions, etc.. Also, we concentrate on cross sectional data. For an overview of semiparametric estimation methods in econometrics we recommend Horowitz (1998), and Härdle, Müller, Sperlich & Werwatz (2004) for a general introduction into these methods.

The rest of the paper is organized as follows. In Section 2 we discuss testing model specification in econometrics, separated in the subsections parametric versus nonparametric, (semi-) parametric versus semiparametric, and non- or semiparametric versus non- or semiparametric models. Section 3 discusses semiparametric estimation, separated in the subsections parametrically specified models with unknown error distribution, structural models with flexible functional forms (with some comments on endogeneity), and unstructured

nonparametric models. Note that this separation is by no means motivated by statistical aspects. It moreover tries to reflect the different tools of methods from the empirical researchers point of view.

2 Testing model specification

2.1 Parametric versus nonparametric models

This was probably the first class of nonparametric tests to verify the specification of econometric models. The null hypothesis consists of a parametric specification of the regression function whereas the alternative is not specified at all to yield an omnibus test. To be more specific: Consider a regression problem $E[Y|X] = m(X)$, $X \in \mathbb{R}^d$, $d \geq 1$, and let $m(\cdot)$ be parametrically specified by m_θ , i.e. a function that is known up to the unknown parameter (vector) θ. Then, the question to test is

$$H_0 : \ m = m_\theta \ \text{ versus } \ H_1 : \ m \neq m_\theta \tag{1}$$

Typical examples are to test the linearity assumption of a simple linear model or to test the link function specification of Probit- and Logit- models.

Even though the following classification is discussable, let us divide the different mathematical approaches for these nonparametric testing problems into the following groups: looking at (integrated) conditional moments, empirical process approaches e.g. combined with Kolmogorov-Smirnov type or Cramer-von-Mises statistics, minimax approaches, and integrated squared differences.

We will not discuss here the differences, advantages and disadvantages of these different approaches but remark one point that could be of interest in practical applications: In the case that the test rejects, the empirical researcher would like to "see" what is this alternative that is considered to be significantly closer to the data generating process (DGP) than its null hypothesis. Many tests of the last mentioned group of tests require an explicit estimation of the alternative. This might be one reason why they are more popular in econometrics. Actually, this last group can be reduced mainly to four different statistics

$$E\left[w_X\{m(X) - m_\theta(X)\}^2\right] \quad , \quad E\left[w_X\{m(X) - m_\theta(X)\}e_X\right] \tag{2}$$

$$E\left[w_X e_X E[e_X|X]\right] \quad , \quad E\left[w_X\{\sigma^2(X) - \sigma^2_\theta(X)\}\right] \ , \tag{3}$$

where e_X is the residuum under the null hypothesis H_0, and w_X a weight function. It is interesting to mention that for finite samples non of these tests has been found to dominate the others, see Dette, von Lieres und Wilkau & Sperlich (2003). Note that almost all tests need resampling methods (usually

wild bootstrap is applied) to find the critical value in practice, i.e. in finite samples.

A main problem with these tests is the SP selection. To circumvent this, recently there is coming up more and more literature on the so called "adaptive testing". The aim is to find a SP that on the one hand holds the wanted first error level and on the other hand maximizes the power of the test.

2.2 (Semi-)parametric versus semiparametric models

Since for practical inference the omnibus tests of Section 2.1 are much too general, apart from the fact that they usually suffer from the curse of dimensionality, there has been developed a class of tests that consider parametric (or semiparametric) null hypotheses versus semiparametric alternatives. This means, only a part of the model is of interest and made more flexible in the alternative. A good example might be to consider generalized (additive) partial linear models of the form $E[Y|X,T] = G\{\beta^T T + \eta(X)\}$, $T^T = (T_1^T, T_2) \in I\!R^q$, $q > 1$, $T_2 \in I\!R^1$, where G is a known link function, β and η unknown. A typical question to test would be

$$\begin{aligned} H_0 &: \quad m(x,t) = G\left\{\beta_1^T t_1 + \beta_2 t_2 + \eta(x)\right\} \quad \text{versus} \\ H_1 &: \quad m(x,t) = G\left\{\beta_1 t_1 + \eta_2(t_2) + \eta(x)\right\} \end{aligned}$$

From a statistical point of view one could just apply (maybe with some minor modifications) the tests statistics introduced in (2) on $\beta_2 t_2$ versus $\eta_2(t_2)$ but this will be very inefficient in many cases.

Additional problems to the ones discussed in Section 2.1 are caused by the nonparametric part in the null hypothesis (in our example $\eta(x)$). Not only that this affects the quality of estimation of both models (null and alternative) and thus the power of the test. Moreover, the necessary resampling methods, in particular wild bootstrap, can be seriously disturbed if the null hypothesis, i.e. the DGP for the bootstrap samples, is poorly estimated.

Example 1.

For 1991, one year after the German unification, we want to investigate the impact of various possible determinants on the intention of East-Germans to migrate to West Germany. The original data set contains 3710 East Germans who were surveyed in 1991 in the Socio-Economic Panel of Germany. Here we consider the data sets from two East German countries: the most northern country of East Germany, i.e. Mecklenburg-Vorpommern (M-V) with $n = 402$, and the most southern one, Sachsen (Sax) with $n = 955$ observations. We use the following variables: family/friend in West, unemployed/job loss certain, middle sized city (10000-100000 habitants) and female [dummies (= 1 if yes, = 0 if no), age (AGE) and household income (HHINCOME) [studentized

continuous variables]. The response is 1 if the person said he is willing to migrate and 0 otherwise.

All methods we use for our study are introduced in Härdle, Huet, Mammen & Sperlich (2003).

In a first step we do a purely parametric logit regression, in a second step we fit a semiparametric generalized additive model for both data sets. Table 1 gives the estimates for the parametric part. For the semiparametric model, we give the results for different SPs, ($h = 1.0$ and 1.25 for M-V, $h = 0.75$, 1.0 for Sax for the directions of interest, $1.1 \cdot h$ for the nuisance directions). In Figure 1 are plotted the additive components for AGE and HHINCOME.

	M-V			Sax		
	par.	semi.a	semi.b	par.	semi.a	semi.b
family/friends West	.5893	.5920	.5809	.7604	.7137	.7289
unemployed/...	.7799	.7771	.7992	.1354	.1469	.1308
middle sized city	.8216	.7156	.7127	.2596	.3134	.2774
female	-.3884	-.3309	-.3485	-.1868	-.1898	-.1871
age	-0.9227	–	–	-0.5051	–	–
hh.income	0.2318	–	–	0.0936	–	–
constant	-1.367	-1.462	-1.411	-1.092	-1.105	-1.101

Table 1: Results of purely parametric estimates (par.) and of the parametric part of a generalized additive partially linear regression model: semi.a (with SP 1.0), semi.b (1.25) for M-V; semi.a (0.75) and semi.b (1.0) for Sax.

The estimates do not depend very much on the chosen bandwidth. Moreover, for the linear part of the model the results are similar to the values of the parametric model. So the qualitative interpretation of the parametric coefficients does not change. In the figure the influence of AGE in M-V does not differ strongly from the influence of AGE in Sax, except that the curve from Sax is more flat in the middle part. In contrast, for HHINCOME the curves from both countries have a totally different shape. On first glance one would guess that AGE could be modeled linearly, at least for M-V. This is less clear for HHINCOME.

In a third step we apply a bootstrap test for linearity to the variables AGE and HHINCOME. We always use 499 bootstrap resamples to determine the critical value. The bandwidths are chosen as above. For the input AGE, linearity is always rejected for the 1 percent level, for all bandwidths in both countries. For the variable HHINCOME, the observed p-values are for M-V .16 [$h = 1.0$], .14 [$h = 1.25$], and for Sax .02 [$h = 0.75$], .01 [$h = 1.0$]. So the deviations for AGE from linearity are more significant. At a first sight, this

seems to be surprising because the plots for HHINCOME differ much more from linearity. Reasons are presumably that the estimates for HHINCOME have large variance and/or the model(s) is (are) misspecified, e.g. the link function $G(\cdot)$ could be misspecified.

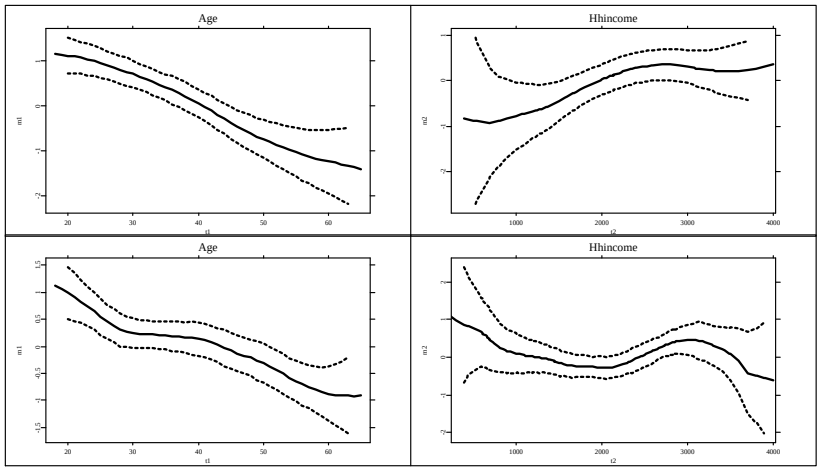

Figure 1: Estimates and 95% uniform confidence bands for the impact of AGE (left) and HHINCOME (right) in M-V with bandwidth 1.25 (upper line), in Sax with bandwidth 0.75 (lower line).

To clarify these two points we construct in a next step uniform confidence bands. In Figure 1, 95% uniform confidence bands are given for the impact functions for M-V. We use SP $h = 1.25$, and 0.75 respectively, always $1.1h$ for the nuisance directions, and $B = 500$ bootstrap replications. All confidence bands contain a linear fit. Only for HHINCOME in Sax the linear fit lies slightly outside the boundary.

In a last step we test the specification of the link function. For testing we use SP $h = 0.75$, ($1.25h$ for nuisance direction) for M-V and $h = 0.6$ for Sax[1]. With $B = 499$ bootstrap replications we get p-values of about 7% for all SPs for M-V and p-values that are always larger than 15% for Sax. So we can conclude that the inconsistency we found in the results for AGE in M-V indeed might be caused by a misspecification of the link $G(\cdot)$. □

[1]For the test further bandwidths are necessary, see Härdle et al (2003). We tried several over a reasonable range and got always very similar results.

2.3 Non- or semiparametric versus non- or semiparametric models

As you can imagine, this title tries to describe something rather general: the check of no parametric specification but of model structures as e.g. additivity, separability, or single index structures. This topic, except additivity testing, can be considered as being still in its infancy.

Here, the maybe most crucial problems are

a) the identification, i.e. the specification of a test statistic that rejects iff H_0 is wrong. E.g. consider weak separable models of the form

$$m(x,t) \;=\; G\{\eta_1(x_1),\ldots,\eta_d(x_d);\theta,t\}\;, \tag{4}$$

G specified up to an unknown parameter θ. Then, if we reject this model, was it because of the weak separability or because of the specification of G ?

b) the choice of the different SPs, in particular under the null model. Now, often the quality of estimating the null hypothesis has a direct effect on the quality of the test in practice. In most cases, if the null model is not estimated sufficiently well, the bootstrap fails completely even though it is consistent, see Dette, von Lieres und Wilkau & Sperlich (2003). Moreover, the SP of the null easily becomes an inherent part of the hypothesis H_0, see also Rodríguez-Poó, Sperlich & Vieu (2002).

3 Non- and semiparametric estimation

3.1 Parametrically specified models with unknown error distribution

A most simple example for estimators in these kind of regression problems are the orthogonal least-squares estimators. For them no new, sophisticated estimation tool is necessary, and they therefore are commonly not mentioned in the context of semiparametric models. But, the hypothesis of unknown error distribution becomes quite a problem when we consider latent variables as response and / or simultaneous equation systems, both rather common in econometrics. Whereas a simple latent variable regression is nothing else than a generalized linear model with unknown link, called single index model (for that we known a huge amount of estimation literature, see Horowitz (1998) or Härdle et al (2004)), the second problem is much more complex:

Consider the selectivity model

$$y = \left\{\beta_0 + x^T\beta_1 + u\right\} d\;, \quad d = \mathbb{1}\{g_\theta(t) > \epsilon\}\;,$$

u, ϵ being error terms, t another vector of explanatory variables, and g_θ a function specified up to θ. Then we can write

$$y = \beta_0 + \beta_1^T x + \lambda\{g_\theta(t)\} + u \ , \qquad \lambda : \ I\!R \to I\!R \text{ smooth}$$

and we want to estimate β_1 (maybe also $\lambda\{g_\theta(t)\}$) semiparametrically. So we do not specify $\lambda(\cdot)$, i.e. the joint distribution of (u, ϵ).

You can apply the so called differencing estimator. For the estimation of β_1, function $\lambda(\cdot)$ is an infinite dimensional nuisance parameter. To get rid of it consider the following difference

$$y_i - y_j = (x_i - x_j)^T \beta_1 + \lambda\{g_\theta(t_i)\} - \lambda\{g_\theta(t_j)\} + u_i - u_j \ , \ i \neq j = 1 \ldots, n.$$

With some weights inverse to $|\lambda(\hat{g}_i) - \lambda(\hat{g}_j)|$, i.e. to $|\hat{g}_i - \hat{g}_j|$ you get

$$\hat{p}_{ij} \ = \ \frac{1}{h} L\left(\frac{\hat{g}_i - \hat{g}_j}{h}\right) d_i d_j \ , \qquad L(\cdot) \text{ some kernel function.}$$

Here, $Z = Z(T)$ are some instruments for X (if needed). The final estimator is

$$\hat{\beta}_1 = \hat{S}_{zx}^{-1} \hat{S}_{zy} \quad \text{with} \quad \hat{S}_{zx} = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \hat{p}_{ij} (z_i - z_j)(x_i - x_j)^T \ .$$

This idea and procedure has been suggested by Powell (1987) but without providing proofs. For them we refer e.g. to Rodríguez-Poó, Sperlich & Fernández (2002). Finally, we would like to remark that this approach has become very popular also in the so called (semiparametric) propensity score analysis.

As these models are essentially parametric, there meet almost none of the criticisms on non- and semiparametric methods (mentioned in the introduction), except the choice of SP and its interpretation. It is clear that the optimal SP is the one that minimizes the mean squared error of $\hat{\beta}_1$. However, it is not that clear how to find this in practice.

3.2 Structural models with flexible functional forms

When we speak of structural models we refer to models that are specified in their structure but not (completely) concerning the functional forms. Typical examples are

a) when the empirical researcher wants to specify his model up to some nuisance parameters; e.g. he includes variables in his model to reduce the noise or avoid endogeneity but does not want to specify its functional impact neither is interested in it.

b) models with some pre-specified separability, additive interaction models, multi index models, etc..

The estimation of (generalized) additive models is already well studied, also the one of additive interaction models (for an overview see Sperlich (1998)), whereas the research on semiparametric estimation of weak and latent separable models (compare equation (4) for its definition) is rather recent, see Rodríguez-Poó et al (2002) and Mammen & Nielsen (2003).

The method of Mammen & Nielsen (2003) is based on smoothed backfitting. On the one hand the identification problems are solved and asymptotic properties developed for a wide range of models, on the other hand the implementation is so far an open problem and it is already clear that the computational expenses will be rather high.

In contrast, Rodríguez-Poó et al (2002) introduce an easy to implement estimation procedure for a wide range of rather general models. However, they could give complete asymptotic theory only for a family fulfilling rather strong (identification) conditions. Their estimation algorithm is based on three-step smoothed likelihood estimation. For identification they need to assume the conditional density of the response as known. They give examples with truncated and censored response variables, in particular the Gronau (1973) model, but allow for flexible functional forms for the η_j, $j = 1, \ldots, d$ (compare equation (4)).

Example 2.

We estimate a female labor supply model for married woman where labor supply is measured in real hours of work. Note that this variable accounts for the number of hours per week the women had declared to work. Many parametric specifications have been tried to model the hours function in this context. A most famous one is the study about the sensitivity against economic and statistic assumptions by Mroz (1987). In our study we only have to specify the error distribution and how we want to combine the nonparametric components. The hours are assumed to be generated by a Tobit 1 model with truncated variables, i.e.

$$y_i = \begin{cases} h(x_i, t_i) + u_i & \text{if } h(x_i, t_i) + u_i > 0,\ u_i \text{ error term} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

We concentrate on a comparison of specifications of possible interactions in h as well as of the behavior of married woman in East and West Germany three years after unification, i.e. in 1993. Those comparisons became quite popular as, due to completely different political, economic and social systems before 1990, the levels of employment of woman where quite different too: in 1993 in the East still about 65%, in the West only about 54%. Consequently, all the studies in the literature have concentrated on participation at all.

We use data taken from the Social Economic Panel of Germany, wave 1993,

cleaned for persons with missing values in the relevant questions and skipping East Germans living in West, West Germans living in East. We have 681 observations for West and 611 for East Germany with a job (i.e. hours > 0). We choose as explaining variables the number of children (Ch1= $\mathbb{1}\{one\ child\}$, Ch2= $\mathbb{1}\{more\ children\}$), education (Edu1= $\mathbb{1}\{high\ school\}$, Edu2= $\mathbb{1}\{academic\ degree\}$) and unemployment rate of the country the person lives in (Urate) for the linear part ($t^T\gamma$). Note that in East Germany there are only 5 countries. For the nonparamertic part $\eta(x)$ we have age of woman (Age), net wage per real hours (Wage), prestige index of their job (PI) and number of years of interruption of professional career (off). For further main income and expenditures we include also the net income of partner per month (Income), and the expenditures for flat minus net income from letting flats (R & L = rent-let). Most probable is an interaction between the determinants of further household income and expenditures apart from the women's one. These are the last two mentioned variables (X_5, X_6). Therefore we study the models of the form

$$\begin{aligned} h_w(t,x) &= t^T\gamma + \eta_1(x_1) + \cdots + \eta_5(x_5) + \eta_5(x_5)\eta_6(x_6) \qquad (6) \\ h_s(t,x) &= t^T\gamma + \eta_1(x_1) + \cdots + \eta_5(x_5) + \eta_6(x_6) \qquad (7) \\ \text{and} \quad h_m(t,x) &= t^T\gamma + \eta_1(x_1) + \cdots + \eta_5(x_5, x_6)\ . \end{aligned}$$

To make them comparable we set $E[\eta_j(x_j)] = 0$, $j = 1, 2, 3, 4, 6$. If by this separability assumption the model is well specified, X_5, X_6 more or less independent, we should get out the same estimates for both specifications, up to a multiplying constant $c = E[\eta_5(x_5)]$ for η_6.

We apply the procedure of Rodríguez-Poó et al (2002). For West Germany we take always SPs $h_j = 1.25\hat{\sigma}_{x_j}$, $j = 1, \ldots, 6$, for East Germany $h_j = 1.5\hat{\sigma}_{x_j}$ as we have less data. Here, $\hat{\sigma}_{x_j}$ indicates the estimated standard deviation of X_j.

Let us first consider the comparison of the different specifications and focus for presentation on the West German data. In Figure 2 and Table 2 (left side for West Germany) we see the results for the additive case h_s. In the table are given additionally the results for a pure parametric linear model (first two columns), all with standard deviations in brackets. In the parametric model we introduced Age**2. This parametric analysis was only done to compare with the parameter estimates $\hat{\theta} = (\hat{\gamma}^T, \hat{\sigma})$ of the semiparametric model. It can be seen that, apart from Edu2 for East Germans, the coefficient estimates do hardly change. But, the error variance (what is not surprising having decreased the degrees of freedom) as well as the variances of the estimates (what is a very good sign) have been reduced a lot using semiparametric methods.

Compare now Figures 2 and 3. In Figure 3 are given the results for the two last component estimates for h_w, $\hat{\eta}_5$ being centered to zero (not for the estimation, only for the presentation). On the bottom of all graphs are

	West Germany				East Germany			
Ch1	-7.847	(1.087)	-6.913	(.7850)	-2.702	(1.054)	-2.152	(.9910)
Ch2	-11.91	(1.221)	-10.84	(.9549)	-2.313	(1.178)	-2.040	(1.130)
Edu1	-.1027	(1.777)	.5738	(1.383)	1.670	(1.300)	1.318	(1.180)
Edu2	.1403	(2.070)	2.125	(2.084)	1.575	(1.610)	4.868	(1.562)
Urate	.2003	(.2254)	.0925	(.1587)	-.5204	(.3242)	-.4256	(.2934)
Age	1.351	(.4662)	-	(-)	1.460	(.4034)	-	(-)
Age**2	-.0184	(1.E-6)	-	(-)	-.0186	(1.E-6)	-	(-)
ln(Wage)	-7.431	(1.067)	-	(-)	-4.126	(.9695)	-	(-)
PI	.2673	(.0436)	-	(-)	.0820	(.0300)	-	(-)
off	-.3485	(.0616)	-	(-)	-.7367	(.1741)	-	(-)
Income	-.1206	(.0245)	-	(-)	-.1200	(.0316)	-	(-)
R & L	.0188	(.0141)	-	(-)	.1092	(.0469)	-	(-)
σ	10.21	(.2961)	6.955	(.1145)	7.828	(.2241)	6.303	(.1803)
Const	24.06	(9.317)	32.44	(-)	30.16	(9.118)	47.25	(-)

Table 2: Results for parametric linear model (columns 1,2 and 5,6) and the semiparametric model (columns 3,4 and 7,8). The standard deviations are given in brackets. In the last line, for the semiparametric model *Const* refers to $\widehat{E}[\eta_5(X_5)] = \frac{1}{N}\sum_i \hat{\eta}_5(x_{i5})$.

given crosses for each observation to indicate the density of the corresponding variable. Up to a multiplying constant c for η_6, they are all the same. For this reason the other components for h_w are not shown as they are exactly the same as we see them in Figure 2. Moreover, c is equal to *Const* from Table 2. This could be taken as an indicator that the model might be well specified by h_s. The estimation of h_m does thus not add any new information.

Now we look on a comparison between the West and the East Germans. As said in the beginning, they come from completely different political, social and economic systems, and though in 1993 at least the political and the economic systems were the same, there were still differences in the economic and political environments. Let us to mention some specials from the East: the unemployment rate was much higher in the East, a higher willingness and motivation of women to search a job, partly based on the lower salaries (compared to the West) of their husbands, a much wider provision of kinder gardens and other possibilities to leave his children. The results are provided in Table 2, Figure 2 (for the West) and 4 (for the East), all based on model h_s. Looking on them, we conclude that behavior for labor supply measured in real hours of work is pretty the same in the East and the West, except for education and number of children. The latter outcome was expected for aforementioned reasons. Comparing this with results of other studies which used the same data base, this is a little bit surprising as they found big

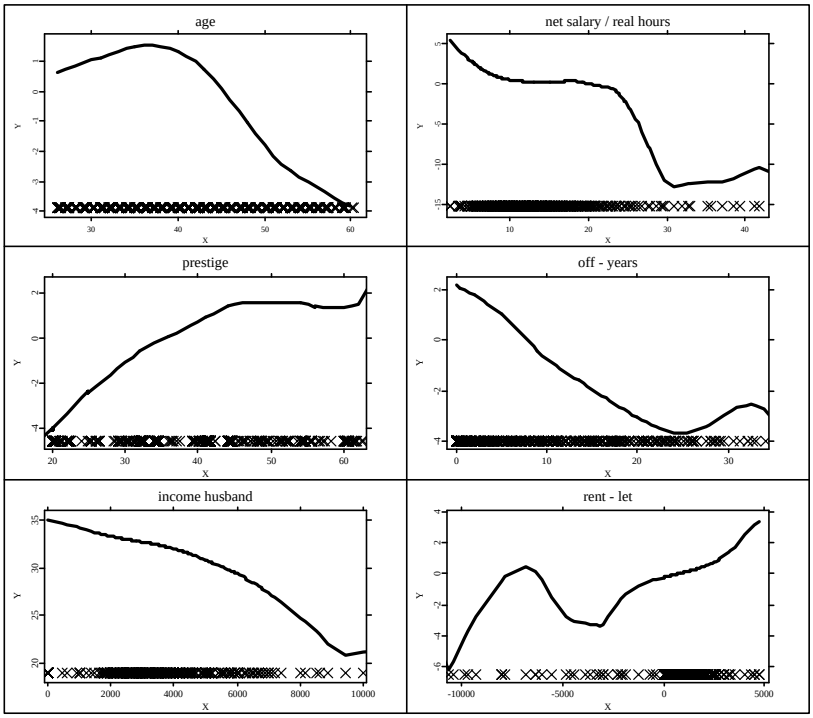

Figure 2: West German women. Results for the additive specification (7). Here, η_5 is centered to zero. Crosses stand for the observations to indicate the density.

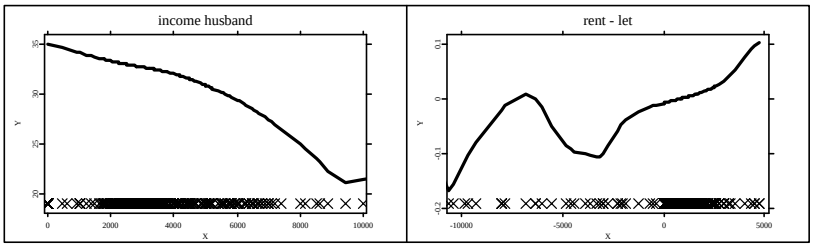

Figure 3: West German women. Results for two last components in specification (6). Here, η_5 is centered to zero.

differences in behavior when looking on participation at all. □

Here now we have faced several of the problems of nonparametric estimation and its solutions (enumerated in the introduction). The lack of the possibility

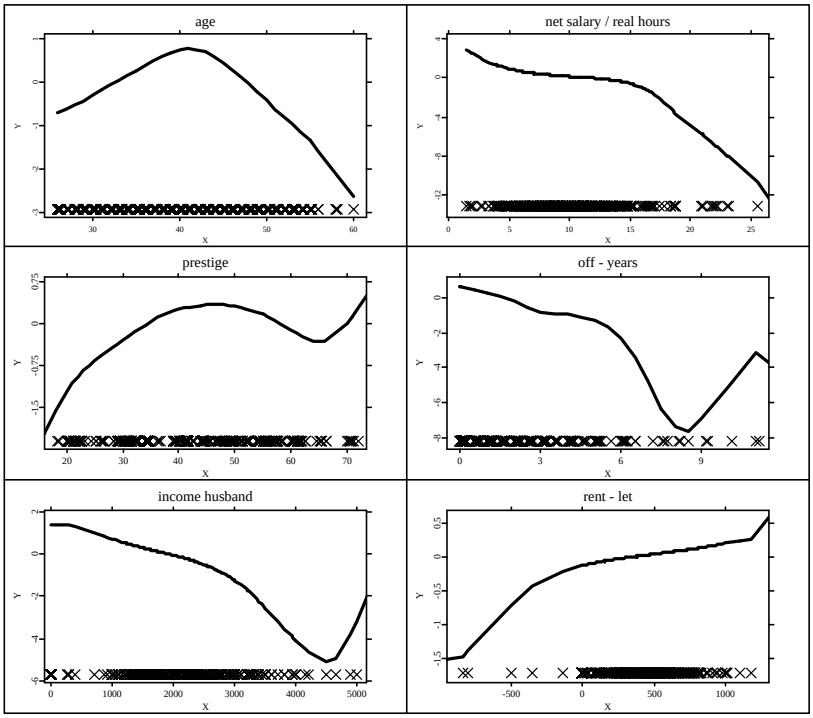

Figure 4: East German women. Results for additive specification (7). Here, η_5 is centered to zero. Crosses stand for the observations to indicate the density.

of modeling has been reduced by semiparametric modeling; also imposing restrictions like monotonicity for the nonparametric part is sometimes possible to impose. This improves automatically the interpretability of the estimates, and often enables or facilitates the estimation of parameters or functions of particular interest (e.g. elasticies, rates of substitution, etc.). Additionally, it can reduce the curse of dimensionality, see Stone (1986), Rodríguez-Poó et al (2002). In those models, the choice of the SP should be considered like choosing the degrees of freedom, i.e. the empirical researcher allows for more flexibility or imposes more smoothness on its functionals. To my opinion, in this context, the "optimality" of the SP has to be defined along the aim of the empirical researcher. Therefore, it is impossible to give here a general rule how to chose it in practice.

Finally let us comment on the problem of endogeneity. To my knowledge, in the context of semiparametric analysis where the regression of interest contains nonparametric functions this problem has been studied first by

Fernández, Rodríguez-Poó, and Sperlich [presented 1999 at ESEM, revised in: Rodríguez-Poó, Sperlich & Fernández (2002)] and by Newey, Powell & Vella (1999). Newey, Powell & Vella (1999) did a profound study on identification. They consider nonparametric (and partially linear) models. Rodríguez-Poó, Sperlich & Fernández (2002) allow for separable and generalized models, therefore they apply assumptions on the error distribution. The two articles coincide in several of the main ideas to circumvent the problem of endogeneity, e.g. they use generated regressors and apply two and / or three step estimators.

3.3 Unstructured nonparametric models

Although "nonparametric" and "unstructured" is essentially the same, we used this title to emphasize the lack of any specification. The only thinkable compromise could be to include partial linear models as long as the linear part serves only to include the impact of dummy variables.

Nonparametric models are useful for optimal prediction (except extrapolation) and explorative data analysis. We might even say: "and for nothing else". The first point is evident because every imposed structure that is not confirmed by the data itself may reduce the quality of the fit. Usually, this approach is interesting whenever we want to predict best whatever the "true model" (if exists) is. Well known examples are financial data problems as predicting stock or bond prices, risks, interest rates, etc.. For a better understanding why and how even totally nonparametric methods can be helpful here see Nielsen & Sperlich (2003).

Less obvious might be the use of nonparametric statistics to explore economic data if the underlying economic process is of interest. To understand this better, let us consider a real data example, taken from Grasshoff, Schwalbach & Sperlich (1999). They do an explorative analysis about the relation of executive pay and corporate financial performance.

Example 3.

Commonly, empirical research concentrates on the pay-performance relationship. Although very different data sets has been adopted the results are always similar showing rather low pay-for-performance elasticities. Almost all studies assume linear or semi-log linear pay functions without applying a test of the adequate functional form. They do not allow for variations across corporations, industries, countries and time. It it assumed that pay functions are homogeneous across corporations, variations are captured by the fixed effects in the constants and assumption about the errors. So it would be interesting to circumvent these possible misspecifications by adopting an explorative data analysis using nonparametric methods. And indeed, the results of Grasshoff, Schwalbach, and Sperlich (1998) show clearly that all

mentioned issues matter, e.g. industry effects are important, assumptions of additivity and linearity are crucial leading to underestimations of the elasticities, etc.. In sum, their results should have far reaching implications for further empirical studies. They also weaken the concern that strong pay-for-performance incentives for executives are missing.

In analyzing executive pay the standard empirical model contains corporate size and financial performance as determinants of pay. Corporate size is a measure of managerial discretion and financial performance is an indicator for managerial incentive compatibility. Both hypotheses are derived from agency theory. Typically, the following regression equation is assumed:

$$\ln C_i^{j,t} = \alpha_{j,t} + \beta^{j,t} P_i^{j,t-1} + \gamma^{j,t} \ln S_i^{j,t-1} + u_i^{j,t} \ , \tag{8}$$

where C_{it} stands for executive pay, P_{it} reflects measures of financial performance and S_{it} represents size for firm i at time t. The terms u_{it} are the stochastic error terms whereas the parameters α_i are mostly modeled as firm-specific fixed effects.

The data base is drawn from varies annual executive pay reports by "Kienbaum Vergütungsberatung". The data contain average annual total pay (fixed and variable) by the top executives of German stock companies (Vorstand of Aktiengesellschaften) and 'companies of limited liabilities' (Geschäftsführer of the Gesellschaft mit beschränkter Haftung). In total, we use data of up to 339 manufacturing firms for the period of 1988 to 1994. Company size is measured by the number of employees and corporate financial performance by the rate of return on sales (ROS). Companies are grouped into the following four distinct industry groups: *(1) Basis industries, (2) Capital goods, (3) Consumer goods* and (4) *Food, drinks and tobacco.* For further details see Grasshoff et al (1999).

To get a primary visual impression of the possible functional forms we first applied the multidimensional, in our case two dimensional, Nadaraya-Watson estimator. The model we estimate is of the form

$$\ln C_i^{j,t} = m^{j,t}\left(P_i^{j,t-1}, \ln S_i^{j,t-1}\right) + u_i^{j,t} \ , \ m^{j,t}: \ \mathbb{R}^2 \to \mathbb{R} \text{ unknown.} \tag{9}$$

We use the quartic kernel with bandwidth $h = 2.5\hat{\sigma}_X$. Notice that since our estimator is a local adaptive one, our results are not effected by possible outliers in the x-direction. For better presentation we show the 3D-plots over trimmed ranges. We have selected the results for two representative years, see Figures 5 and 6. Considering the plots over the years we can realize strong functional similarities between the industry groups 1 and 2 while the results for the other groups seem not to be homogeneous at all. Regardless the outliers we see a strong positive relation for compensation to firm size at least for group 1 and 2, and a weaker one to the performance measure varying over years and groups.

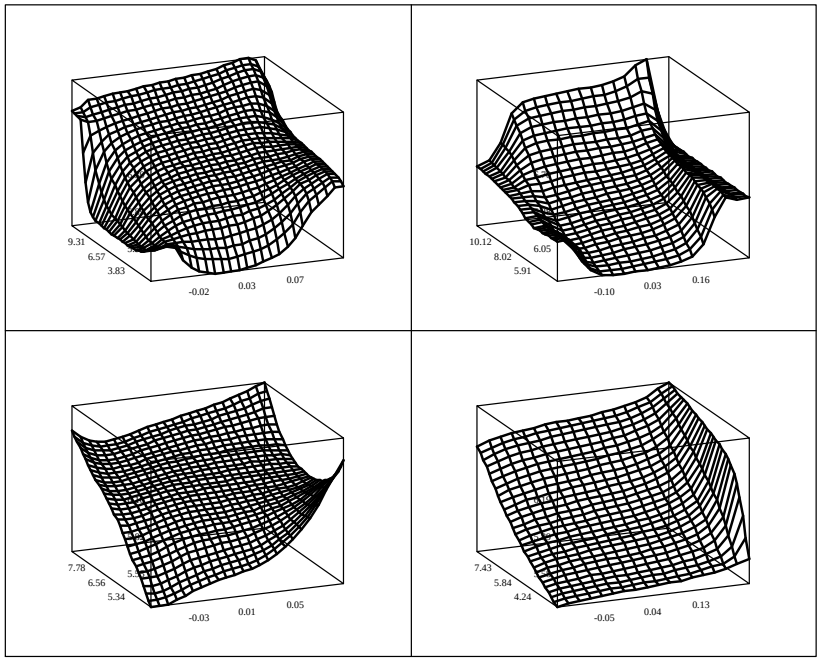

Figure 5: The 2-dimensional Nadaraya-Watson estimation for 1989/90. Plotted are the expected executive pay on size (left axes) and ROS (right axes). First row: group1 and 2, Second row: group 3 and 4.

Further we can recognize some interaction of the independent variables especially in group 3 and 4. This can visually be detected as follows. Imagine you cut slices parallel to the x-axes. If these slices indicate different functional forms within one direction separability of the inputs is not justified. Regarding this procedure we state additivity for group 1 and 2.

Next, a study was done where the regression function was modeled additively with backfitting. This study is skipped here as it did not yield much significant new insight. For more details see Grasshoff et al (1999).

Finally, we estimate the pure marginal effects of the independent variables. The model we estimate can be imagined as general as (9), but we only estimate the marginal effects, not the joint regression function $m(\cdot)$ (skipping the indices (j,t) above). If the model is of additive form $m(x_1,x_2) = m_1(x_1)+m_2(x_2)$, then the marginal effects correspond to m_1, m_2. We use a local linear kernel smoother with quartic kernel and bandwidths $h = 1.5\hat{\sigma}_{x_k}$, $k = 1,2$, ($2.5\hat{\sigma}_{x_k}$ for the nuisance directions). We present the estimation results together with confidence intervals in forms of $2\hat{\sigma}(\hat{m}_k(x_k))$ -bands, where $\hat{\sigma}(\hat{m}_k(x_k))$ indicates the estimated standard deviation of additive component

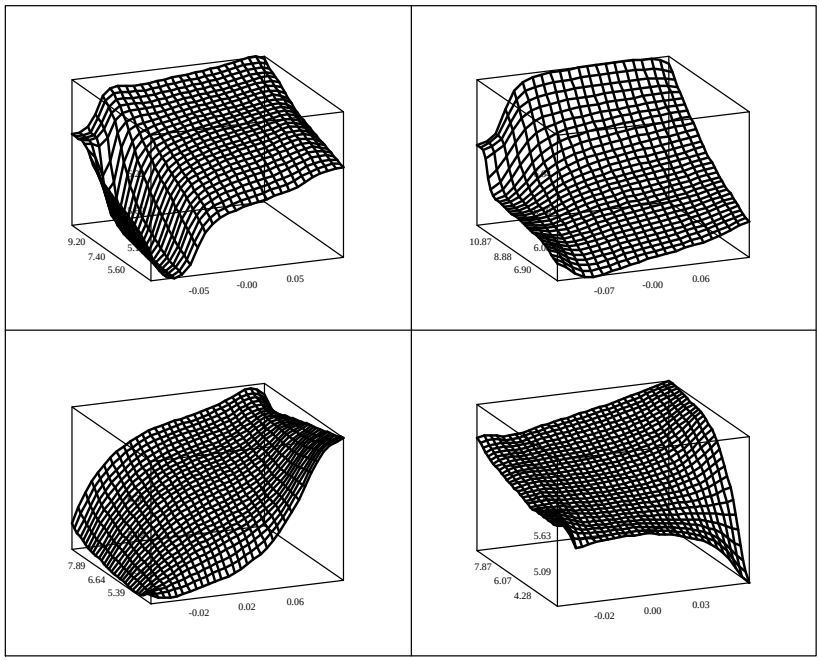

Figure 6: The 2-dimensional Nadaraya-Watson estimation for 1990/91. Plotted are the expected executive pay on size (left axes) and ROS (right axes). First row: group 1 and 2, Second row: group 3 and 4.

$\hat{m}_k$ at point x_k.

As a main result we can postulate that these estimation results are consistent with the findings above. First, the nonlinearities of the financial performance influence are strengthened especially for groups 1 and 2. Second, it seems that interactions are present, so the assumption of additivity would be wrong what renders an economic interpretation rather difficult. □

Here now meet more ore less all the criticisms against nonparametric methods mentioned in the introduction. Interpretability is hardly given, neither for the estimates nor for the SP choice. In general, the Sps should be chosen data driven (e.g. by cross validation or plug in) to minimize the estimation error or optimize prediction power were prediction is pretended. Certainly, interpretability of the estimates is not necessary if one only wants to predict, but it is of interest if one wants to make studies as in our example. Further, since the curse of dimensionality kicks in rapidly, the possibilities of these methods are rather limited unless you have really large samples. E.g. in our example, a test for additivity would simply not work, see Dette, von Lieres und Wilkau & Sperlich (2003).

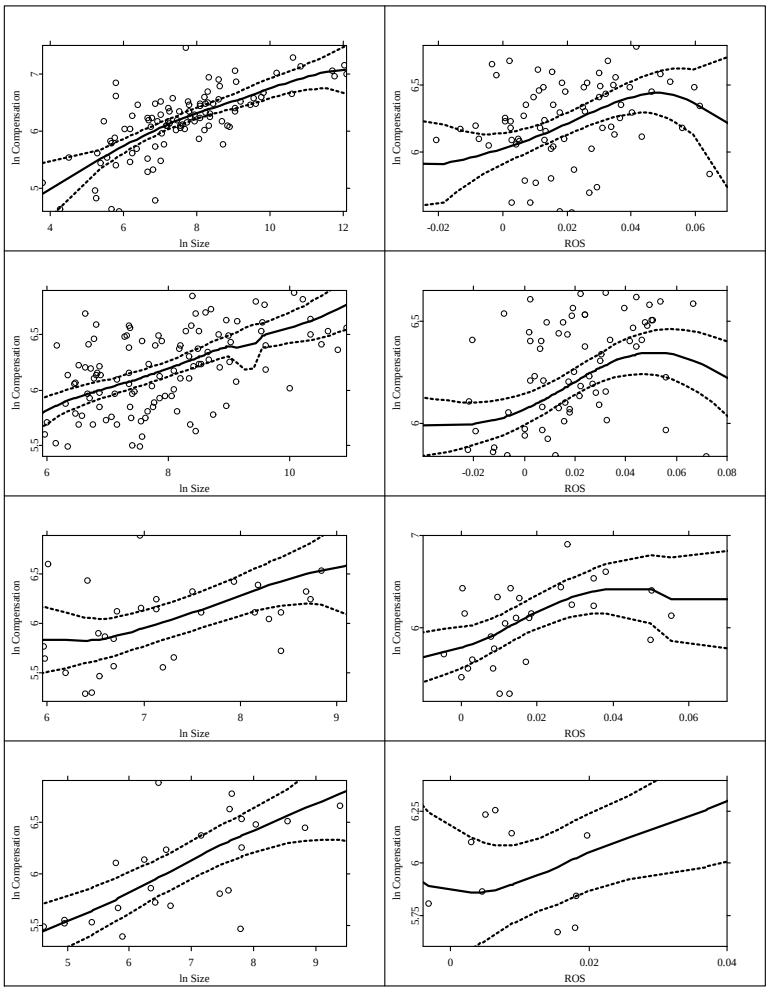

Figure 7: Marginal Integration estimates for 1991/92 with $2\hat{\sigma}(\hat{m}_k(x_k))$, $k = 1, 2$ bands for industry groups 1-4 from top to bottom.

References

Dette, H., von Lieres und Wilkau, C. & Sperlich, S. (2003), 'A comparison of different nonparametric methods for inference on additive models, *Nonparametric Statistics*, forthcoming

Grasshoff, U., Schwalbach, J. & Sperlich, S. (1999), 'Executive Pay and Corporate Financial Performance: an Explorative Data Analysis, *Work-*

ing paper 99-84 (33), Universidad Carlos III de Madrid

Gronau, R. (1973), 'The Effects of Children on the Housewife's Value of Time. *Journal of Political Economy* **81** , 168–S199.

Härdle, W., Müller, M., Sperlich, S. & Werwatz, A. (2004), *Non - and Semiparametric Modelling*, to appear in Springer Verlag

Härdle, W., Huet, S., Mammen, E. & Sperlich, S. (2003), 'Bootstrap Inference in Semiparametric Generalized Additive Models, *Econometric Theory*, in press

Horowitz, J. (1998). *Semiparametric Methods in Econometrics*, Springer.

Mroz, T.A. (1987), 'The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions, *Econometrica* **55** (4), 765–799.

Newey, W.K., Powell, J.L. & Vella, F. (1999), 'Nonparametric Estimation of Triangular Simultaneous Equation Models, *Econometrica* **67** (3), 565–604.

Mammen, E. & Nielsen, J.P. (2003), 'Generalised Structured Models, *Biometrika*, in press

Nielsen, J.P. & Sperlich, S. (2003), 'Prediction of stocks: A new way to look at it, *Astin Bulletin* **33.2**, in press

Powell, J. L. (1987), 'Semiparametric Estimation of Bivariate Latent Variable Models, *Working Paper, University of Wisconsin - Madison*

Rodríguez-Poó, J.M., Sperlich, S. & Fernández, A.I. (2002), 'Semiparametric Three Step Estimation Methods for Simultaneous Equation Systems, *Working Paper, Carlos III de Madrid*

Rodríguez-Poó, J.M., Sperlich, S. & Vieu, P. (2003), 'Semiparametric Estimation of Separable Models with Possibly Limited Dependent Variables, *Econometric Theory*, in press

Sperlich, S. (1998), *Additive Modelling and Testing Model Specification*, Shaker Verlag, Aachen.

Stone, C. J. (1985), 'Additive regression and other nonparametric models, *Annals of Statistics* **13**(2), 689–705.

Stone, C. J. (1986), 'The dimensionality reduction principle for generalized additive models, *Annals of Statistics* **14**(2), 590–606.