

This article was downloaded by: [National Chiao Tung University 國立交通大學]

On: 28 April 2014, At: 06:35

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Communications in Statistics - Theory and Methods

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/lsta20>

Tensor product polynomial splines

Lin-An Chen^a, Tzong-Shi Lee^a & Wenyaw Chan^b

^a Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan

^b School of Public Health, University of Texas-Houston, Houston, Texas

Published online: 05 Jul 2007.

To cite this article: Lin-An Chen, Tzong-Shi Lee & Wenyaw Chan (1997) Tensor product polynomial splines, Communications in Statistics - Theory and Methods, 26:9, 2093-2111, DOI: [10.1080/03610929708832036](https://doi.org/10.1080/03610929708832036)

To link to this article: <http://dx.doi.org/10.1080/03610929708832036>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Tensor Product Polynomial Splines

Lin-An Chen*, Wenyaw Chan and Tzong-Shi Lee*

* Institute of Statistics, National Chiao Tung University,
Hsinchu, Taiwan.

School of Public Health, University of Texas-Houston,
Houston, Texas.

Keywords: Nonparametric regression; Piecewise polynomials; Tensor product spline.

Abstract

This paper presents a unified study of smoothing tensor product piecewise polynomials as tensor product polynomial splines. This study generalizes the theory of univariate polynomial splines of Poirier (1973) and Smith (1979) to the multivariate setting, and the method proposed herein provides various spline bases that gives great flexibility for selecting models. An example analyzing an Australia wine industry data shows simple tensor product spline can capture the trend in the data appropriately to law of economics whereas multiple linear model fails in capturing the regression trend (see Maddala (1988)).

1. Introduction

In estimation of the mean response function μ for the following nonparametric regression model

$$y = \mu(x) + \epsilon,$$

the polynomial spline (also called least squares spline or regression spline) has played a prominent role in nonparametric regression. The polynomial spline is a piecewise polynomial; each defined over a different subregion of the domain, constrained to have continuous lower order derivatives or partial derivatives for the multivariate case as well. Representatives of univariate polynomial splines are introduced by Poirier (1973), Buse and Lim (1977) and Smith (1979). Among these, Smith (1979) demonstrated that the set of univariate polynomial splines (order k) forms a vector space spanned by a basis including polynomial terms and truncated “+” functions (see also de Boor (1978) and Wegman and Wright (1983)) as follows:

$$x^{c_0}, (x - \delta_1)_+^{c_1}, \dots, (x - \delta_a)_+^{c_a}, c_i = 0, 1, \dots, k \text{ and } i = 0, 1, \dots, a \quad (1.1)$$

where $\{\delta_1 < \delta_2 < \dots < \delta_a\}$ represents the mesh of knots and $a > 0$ is the number of knots. This representation is clearly a very useful way since it casts the piecewise polynomial problem into a linear model context. Then its coefficients may be estimated by least squares method and many of the F - and t - procedures can be applied to determine whether the coefficients are zeros. Although other choices can be made for basis formulation, the representation of (1.1) is also meaningful for polynomial spline functioning in the role of an approximation tool. From Taylor theory, linear function of x^{c_0} 's represents a polynomial approximation of μ and linear combination of truncated functions then represents an approximation of the residual caused from polynomial approximation (see Eubank (1988) for this point). With this representation, we can test the hypothesis of μ being a polynomial or even a simple linear regression function.

In estimating multivariate mean response function μ (x is a vector), linear combinations of tensor products of truncated spline functions have recently been considered as candidate set for model selection. With such a candidate set, Friedman and Silverman (1989), Friedman (1991) and Zhang (1994)

proposed various algorithms for effectively determining the spline model, so-called adaptive spline. In their studies, the first order truncated “+” and “-” spline functions and their tensor products are served as the candidate set. Reflecting the point indicated by Breiman (1991) that the adaptive splines will become widely used in application, Hastie, Tibshirani and Andreas (1994) applied it to obtain nonparametric versions of discriminant analysis, and LeBlanc and Tibshirani (1994) used it to develop a nonlinear generalization of principal component analysis.

Although the tensor product splines is a feasible scheme in statistical application, its theory of being a smoothing piecewise multivariate polynomials has not been investigated. As Eubank (1988, p371) pointed out, the tensor product technique requires further research before its use can be a standard method of curve fitting. With this need, this paper presents unified study of the tensor product splines. In this study, bases of tensor product splines with various smoothness conditions are derived. Representations of tensor product spline provided in this paper all allow a test for the hypothesis of μ being a multivariate polynomial regression model or even a multiple linear regression model whereas this test is not allowed to do with the adaptive spline due to its representation. Moreover, admitting choices of polynomial order and smoothness conditions makes the tensor product spline flexible for model selection. Through analyzing a data set of Australia wine industry we also see that a simple tensor product spline model can successfully capture the trend in the data, appropriate to law of economics. In this data analysis, we introduce a robust estimation-based technique for knots determination and setting of region decomposition. Notably, tensor product splines have also been used in multivariate function estimation by Stone (1994). Some multivariate smoothers are available in statistical literature other than the tensor product splines. For instance, a bivariate polynomial spline by Chen (1996) that formulated a spline as a linear restricted least squares estimate, the ATS method by Cleveland, Mallows, and McRae (1993), the locally weighted regression method by Cleveland and Devlin (1988), and the thin plate method by Wahba (1979) and Gu and Wahba (1993).

We will proceed as follows. In Section 2 we introduce the class of tensor product piecewise polynomials and state two equivalent formulations of these piecewise polynomials that helps the derivation of spline bases. Bases of tensor product piecewise polynomials and those of tensor product splines are then stated in Section 3. Section 4 provides a data analysis of the Australia wine data. Finally, the proofs of lemmas and theorems are given in Section 5.

2. Tensor Product Piecewise Polynomials

Suppose that we have explanatory variables x_1, \dots, x_p and meshes of knots $\{\delta_1^1 < \dots < \delta_{a_1}^1\}, \dots, \{\delta_1^p < \dots < \delta_{a_p}^p\}$. Let $\delta_0^i < \delta_1^i$ and $\delta_{a_i+1}^i > \delta_{a_i}^i, i = 1, \dots, p$ such that the p -dimensional rectangle $\{(x_1, \dots, x_p) : \delta_0^i < x_i < \delta_{a_i+1}^i, i = 1, \dots, p\}$ is the domain of the regression function. A tensor product polynomial is defined as

$$p = \sum_{c_p=0}^k \dots \sum_{c_1=0}^k \beta_{c_1 \dots c_p} x_1^{c_1} \dots x_p^{c_p}$$

and a tensor product piecewise polynomial is of the form

$$\sum_{t_p=0}^{a_p} \dots \sum_{t_1=0}^{a_1} p^{t_1 \dots t_p} I(\delta_{t_i}^i < x_i \leq \delta_{t_i+1}^i; i = 1, \dots, p) \quad (2.1)$$

where $p^{t_1 \dots t_p}$ are tensor product polynomials. A tensor product piecewise polynomial regression model is a function in (2.1) plus an error variable. This setting generalizes many linear econometric models to piecewise models. In one example we will study later, some explanatory variables are correlated to the error term. Clearly, the space of tensor product piecewise polynomials has dimension $(k+1)^p \prod_{i=1}^p (a_i+1)$. Herein, we define the tensor product splines of this type with uniform smoothness condition in its partial derivatives. Other types of smoothness conditions can also be imposed on the tensor product piecewise polynomials to form various tensor product splines.

Definition 2.1. For $0 \leq j \leq k$, a tensor product piecewise polynomial is

called a smoothness- j tensor product spline if its $(j_1 \dots j_p)$ -th partial derivative is continuous, which holds for $0 \leq j_1 + \dots + j_p \leq j$.

As seen, a smoothness- k tensor product spline must have coefficients of the term $x_1^{c_1} \dots x_p^{c_p}$ that correspond to the polynomial pieces in all cells all equal. This statement also holds for all $(c_1 \dots c_p)$. Then, this spline turns out to be a multivariate polynomial. In the rest of this paper, we assume that $0 \leq j \leq k - 1$. With this setting, the tensor product spline is smooth as at least a continuous function ($j = 0$). The adaptive splines used in Friedman (1991) are part of this spline space.

The spline bases derived herein are motivated by the following proposition that provide an ANOVA-like decomposition of tensor product piecewise polynomials.

Proposition 2.2. A tensor product piecewise polynomial of (2.1) is equivalent to the following two formulations: (a)

$$\begin{aligned}
 & p^{0 \dots 0} + \sum_{s=1}^p \sum_{t_s=1}^{a_s} p^{0 \dots t_s \dots 0} I(x_s > \delta_{t_s}^s) \\
 & + \sum_{1 \leq s_1 < s_2 \leq p} \sum_{t_{s_2}=1}^{a_{s_2}} \sum_{t_{s_1}=1}^{a_{s_1}} p^{0 \dots t_{s_1} \dots t_{s_2} \dots 0} I(x_{s_1} > \delta_{t_{s_1}}^{s_1}, x_{s_2} > \delta_{t_{s_2}}^{s_2}) \\
 & \quad \vdots \\
 & + \sum_{t_p=1}^{a_p} \dots \sum_{t_1=1}^{a_1} p^{t_1 \dots t_p} I(x_1 > \delta_{t_1}^1, \dots, x_p > \delta_{t_p}^p)
 \end{aligned}$$

where all p^{\dots} are tensor product polynomials.

(b) the piecewise polynomial of (a) with replacing all indices “>” in indicator function $I(\)$ by “<”.

The dimension of the space of tensor product piecewise polynomials in (2.1)

remains to be true for the space in (a) or (b) of the above proposition. It is meaningful in the above regression function decomposition that the polynomial $p^{0 \dots 0}$ demonstrates the main regression trend whereas the truncated polynomials demonstrate the block of rectangular region interactions. Without forcing smoothness conditions, this decomposition provides a relatively easy test for assuming a polynomial regression through usual t - or F -test when the variables x_1, \dots, x_p are independent of the error variable. These properties still hold for the tensor product splines.

3. Tensor Product Polynomial Splines

Denote “+” function by $x_+ = \max\{0, x\}$ and “-” function by $x_- = x - x_+$. For convenience, we state the set of truncated functions:

$$\begin{aligned} & \prod_{i=1}^p x_i^{c_i}, \\ & \prod_{i \neq s} x_i^{c_i} (x_s - \delta_{t_s}^s)_{*}^{c_s}, t_s = 1, \dots, a_s \text{ and } s = 1, \dots, p, \\ & \prod_{i \neq s_1, s_2} x_i^{c_i} (x_{s_1} - \delta_{t_{s_1}}^{s_1})_{*}^{c_{s_1}} (x_{s_2} - \delta_{t_{s_2}}^{s_2})_{*}^{c_{s_2}}, t_{s_1} = 1, \dots, a_{s_1} \text{ and } t_{s_2} = 1, \dots, a_{s_2}, \\ & \text{for } 1 \leq s_1 < s_2 \leq p, \\ & \vdots \\ & \prod_{s=1}^p (x_s - \delta_{t_s}^s)_{*}^{c_s}, t_s = 1, \dots, a_s. \end{aligned} \tag{3.1}$$

The above are “+” functions or “-” functions depending on setting “*” as “+” or “-” respectively. We then state the theory providing bases for the tensor product piecewise polynomials.

Theorem 3.1. The functions in (3.1) with $* = +$ or $* = -$ and where c_i, c_s, c_{s_1} and $c_{s_2} = 0, 1, \dots, k$ are both bases of the space of tensor product piecewise polynomials.

Each basis has elements of number exactly $(k+1)^p \prod_{i=1}^p (a_i + 1)$. We then know that a tensor product polynomial spline of any smoothness conditions

defined on partial derivatives is a linear function of the elements in basis either the one with “+” function or the one with “-” function. We now state the main results of representations for the uniform smoothness- j tensor product splines.

Theorem 3.2. The functions in (3.1) with $\star = +$ or $\star = -$ and where $c_i = 0, 1, \dots, k$, and $c_{s_1}, c_{s_2} = j + 1, \dots, k$ are both bases of the space of smoothness- j tensor product polynomial splines.

The dimension of smoothness- j tensor product splines is $(k - j)^p \prod_{i=1}^p (a_i + 1)$. The smoothest tensor product polynomial spline is the one with $j = k - 1$ which can be written as a linear combination of the following functions either with “ $\star = +$ ” or “ $\star = -$ ”:

$$\prod_{i=1}^p x_i^{c_i}, \prod_{i \neq s} x_i^{c_i} (x_s - \delta_{t_s}^s)_\star^k, \prod_{i \neq s_1, s_2} x_i^{c_i} (x_{s_1} - \delta_{t_{s_1}}^{s_1})_\star^k (x_{s_2} - \delta_{t_{s_2}}^{s_2})_\star^k, \dots, \prod_{s=1}^p (x_s - \delta_{t_s}^s)_\star^k.$$

The tensor product spline considered by Friedman (1991) and Zhang (1994) is, with $\star = +$ and $-$, formulated as linear combination of the following truncated functions in addition with constant 1:

$$\begin{aligned} &1, (x_s - \delta_{t_s}^s)_\star, s = 1, \dots, p, \\ &(x_{s_1} - \delta_{t_{s_1}}^{s_1})_\star (x_{s_2} - \delta_{t_{s_2}}^{s_2})_\star, 1 \leq s_1 < s_2 \leq p, \\ &\vdots \\ &\prod_{s=1}^p (x_s - \delta_{t_s}^s)_\star. \end{aligned} \tag{3.2}$$

The above functions with \star fixed as $+$ or $-$ form a linearly independent set with elements of a number $\prod_{i=1}^p (a_i + 1)$. This set contains continuous piecewise polynomials of order $k = 1$, it is also a generator of smoothness-0 tensor product spline of order 1. However, the candidate set adopted by Friedman (1991) and Zhang (1994) includes functions of (3.2) with both $+$ and $-$ functions which is large as twice the number $\prod_{i=1}^p (a_i + 1)$. Breiman

(1991) pointed out this fact of redundant candidate set. Moreover, no obvious way is available to generalize the basis functions of (3.2) to the basis of smoothness- j tensor product spline of arbitrary order k . The tensor product splines then provide greater flexibility for model selection. On the other hand, a tensor product spline model is able for us to test the hypothesis either a multivariate polynomial regression function

$$H_0 : \mu(x_1, \dots, x_p) = \sum_{c_p=0}^k \dots \sum_{c_1=0}^k \beta_{c_1, \dots, c_p} \prod_{i=1}^p x_i^{c_i}$$

or a multiple linear regression

$$H_0 : \mu(x_1, \dots, x_p) = \sum_{i=1}^p \beta_i x_i$$

through the usual t - or F -test procedures when x_1, \dots, x_p are independent variables whereas these tests are not obviously able applying to the adaptive splines.

In application, a problem occurred quite often in structural change model is that the regression function switch in a multivariate rectangular region, a subset of the domain of some explanatory variables. The following using truncated “+” and “-” functions formulates a simple switch regression model:

$$y = \beta_0 + \sum_{i=1}^p \beta_i x_i + \sum_{i \in S} \gamma_i (x_i - \delta_1^i)_+^k (x_i - \delta_2^i)_-^k + \epsilon \quad (3.4)$$

where $\delta_m^i, m = 1, 2$ are switch points and S is a switch index set.

The switch regression appears in (3.4) is a tensor product spline forcing equality of polynomial coefficients those lie outside the region

$$\{(x_1, \dots, x_p) : \delta_1^i < x_i < \delta_2^i \text{ for } i \in S \text{ and } x_i \in R \text{ for } i \notin S\}$$

Model (3.4) will be used in next section analyzing a data set of wine industry.

4. Example of Fitting Spline Through Robust Estimation

Consider a data with two dependent variables of wine industry from year 55-56 to 74-75 (1955-1956 to 1974-1975) in Australia (see Maddala (1988)).

For abbreviation, when we say year n that means from year n to year $n + 1$. The unknown demand model in wine industry is

$$y_1 = \mu(y_2, z_1, z_2, z_3, \beta) + \epsilon \tag{4.1}$$

where all variables are in logs. A linear model with more than one dependent variable is called the linear simultaneous equations model in econometrics (see for example, Theil (1971)). In this model, y_1 and y_2 are dependent variables representing the consumption and price of wine, respectively. Also, $z_i, i = 1, 2$ and 3 , are independent variables representing the price of beer, disposable income, and advertising expenditure. From law of economics, we anticipate values of parameters β making the regression function μ nonincreasing in variable y_2 and nondecreasing in variables z_1, z_2 and z_3 . There is an available instrumental variable z_4 , the index of storage costs. Moreover, the dependent variables y_1 and y_2 are reasonably formulated as a bivariate linear regression model in terms of independent variables $z_i, i = 1, \dots, 4$. For linear simultaneous equations model, a usual way in estimating regression parameters β is firstly estimating parameters of the bivariate regression model and then replace y_2 by its predictor \hat{y}_2 in the linear model μ and secondly computing estimates of β from this new linear model with y substituted by \hat{y} . This is so-called two stage estimation. The need for this problem is to select a suitable regression function that can appropriately reflect the relationship between these economic variables. We then want to show that a simple tensor product spline fits this need.

A linear mean response μ as

$$y_1 = \beta_0 + \beta_{pw}y_2 + \beta_{pb}z_1 + \beta_dz_2 + \beta_ez_3 + \epsilon \tag{4.2}$$

has been selected to fit the regression model (4.1) by Maddala (1988), where, by law of economics, we anticipate < 0 for β_{pw} and > 0 for the other parameters. Maddala (1988) applied a classical estimation technique (two stage least squares estimation) to estimate the parameters. This is a least

squares analogue to the simultaneous equation model. However, besides β_d all the other parameters are estimated in the wrong signs. This result provides us a suspicion that using whole data set to fit a linear model is not appropriate.

For obtaining a tensor product spline to fit this data set, we rely on robust estimation method to detect the knots and the main trend of the regression function. Let the estimate of y_{2i} be

$$\hat{y}_{2i} = \hat{\pi}_0 + \hat{\pi}_1 z_{1i} + \hat{\pi}_2 z_{2i} + \hat{\pi}_3 z_{3i} + \hat{\pi}_4 z_{4i}$$

with values $\hat{\pi}_i, i = 0, 1, \dots, 4$ solving

$$\min_{\pi_i, i=0,1,\dots,4} \sum_{i=1}^n |y_{2i} - (\pi_0 + \pi_1 z_{1i} + \dots + \pi_4 z_{4i})| \quad (4.3)$$

(see Koenker and d'Orey (1987) for computation). This defines the first stage estimation of dependent variable y_2 . For the rest of this section, all second stage estimations are based on the following model

$$y_{1i} = \mu(\hat{y}_{2i}, z_{1i}, z_{2i}, z_{3i}, \beta) + \epsilon^*. \quad (4.4)$$

We unify the first stage estimation for comparisons standing on a common ground.

The following table lists the ℓ_1 -norm estimates.

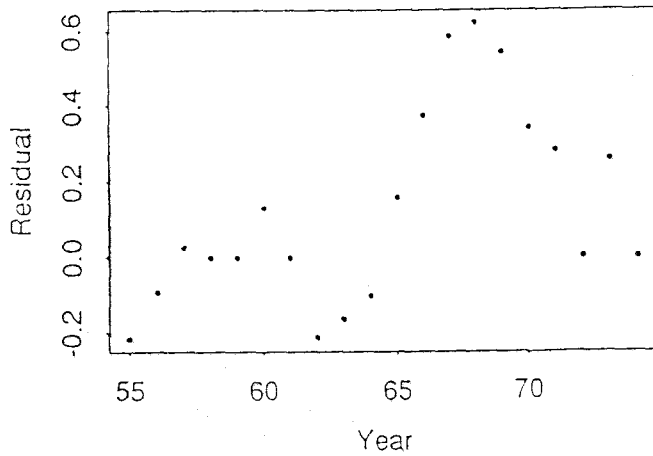
The signs of parameters corresponding to regressors y_2, z_1, z_2 and z_3 are all in the right signs. However, the ℓ_1 -norm is more or less estimating a phase of regression function. To see this, the plot of the residuals computed from the ℓ_1 -norm estimates in the following figure shows the suspicion that there exists structural change of regression function for the observations more or less in years 66 to 71.

Since each regressor variable has observations for years 66 to 71 falling in an interval, it is then reasonable to select regression function from the following model class,

Table 1. Two stage ℓ_1 -norm estimates

param.*	β_0	β_{pw}	β_{pb}	β_d	β_e
ℓ_1	-25.87	-6.19	1.868	2.642	.830

* param.=parameter



$$y_1 = \beta_0 + \beta_{pw}y_2 + \beta_{pb}z_1 + \beta_dz_2 + \beta_ez_3 + \gamma_1(y_2 - a_1)_+(y_2 - a_2)_- \quad (4.5)$$

$$+ \gamma_2(z_1 - b_1)_+(z_1 - b_2)_- + \gamma_3(z_2 - c_1)_+(z_2 - c_2)_- + \gamma_4(z_3 - d_1)_+(z_3 - d_2)_- + \epsilon.$$

In this spline model, the linear function $\beta_0 + \beta_{pw}y_1 + \dots + \beta_ez_3$ demonstrates the main trend regression and the truncated functions provide showing the interactions of explanatory variables in the multi-rectangular region.

For estimating knot points, we want to determine the years for which their corresponding observations of regressors falling in the multi-rectangular region, the domain of the truncated functions. This is equivalent to determine those years for which their corresponding observations of variable y_1 completely follow the main trend regression function.

With observations removed in year period (as indicated in Table 2), we compute least squares estimate based on model (4.4) with μ linear as it in

Table 2. Appropriateness and SSR for year period removing

years removed	SSR	approp.*	years removed	SSR	approp.
66-71	.145	yes	61-71	.106	no
65-71	.114	yes	60-71	.074	no
64-71	.115	yes	60-72	.049	no
63-71	.114	yes	60-73	.176	yes
62-71	.094	no	60-74	.162	no

* approp.=appropriateness

(4.2), where the year period removing is determined based on the absolute residuals through the ℓ_1 -norm estimation shown in Figure 1. The reason that we still fit a linear model is to find the subsets of data that can be explained by the main trend regression (linear function). The results of SSR and status of appropriateness in sign are listed in Table 2.

Again, although linear model is not suitable in explaining the whole data set, however, some subsets of the data are valid to do that way. There are 5 ways of years removing so that the least squares technique can obtain estimates appropriate in sign. We select the domain of explanatory variables for year period 65 to 71 as the second trend regression region since its period is shorter and its corresponding SSR is the smallest. This determine the estimates of knots. The knot candidates shown from the observations of regressors is stated in the following table.

Plugging these knot values into regression function in (4.5), we will determine a sub-model that can appropriately explain the whole set of data (truncated functions are imposed into the model at this step). We compute the least squares estimates based on the whole data set for all candidate models, specified by parameters γ_i associated with truncated functions, and give the SSR and appropriateness of each model in Table 4 (SSR are not listed for inappropriate cases).

Table 3. Knot candidates

a_1	a_2	b_1	b_2	c_1	c_2	d_1	d_2
4.60	4.90	3.53	3.61	7.12	7.35	4.60	5.0

Table 4. Comparisons of submodels

truncated function	years in second trend	appropriate.	SSR
None	None	No	*
γ_1	65 - 74	No	*
γ_2	65 - 72	No	*
γ_3	64 - 71	No	*
γ_4	63 - 71	No	*
γ_1, γ_2	65 - 72	No	*
γ_1, γ_3	65 - 71	yes	.175
γ_1, γ_4	65 - 71	No	*
γ_2, γ_3	65 - 71	Yes	.180
γ_2, γ_4	65 - 71	Yes	.153
γ_3, γ_4	64 - 71	Yes	.255
$\gamma_1, \gamma_2, \gamma_3$	65 - 71	Yes	.167
$\gamma_1, \gamma_2, \gamma_4$	65 - 71	Yes	.151
$\gamma_1, \gamma_3, \gamma_4$	65 - 71	Yes	.172
$\gamma_2, \gamma_3, \gamma_4$	65 - 71	Yes	.152
$\gamma_1, \gamma_2, \gamma_3, \gamma_4$	65 - 71	Yes	.151

The submodel with truncated functions associated with parameters γ_2 and γ_4 is with SSR closer to its smallest value and we set it as the estimate of tensor product spline. The reason is that, although the one with truncated functions associated with parameters γ_1, γ_2 and γ_4 has smaller SSR, however, adding one truncated function only reduce SSR with .001. The selected spline model is

$$y_1 = \beta_0 + \beta_{pw}y_2 + \beta_{pb}z_1 + \beta_dz_2 + \beta_ez_3 + \gamma_1(z_1 - 3.53)_+(z_1 - 3.61)_- + \gamma_2(z_3 - 4.60)_+(z_3 - 5.0)_- + \epsilon$$

Table 5. ℓ_1 -norm estimates

parameter	β_0	β_{pw}	β_{pb}	β_d	β_e	γ_1	γ_2
estimate	-23.03	-.37	.45	3.22	17	-242.4	-5.47

and the least squares estimates of these parameters are listed in the following table.

The parameter estimate $(\hat{\beta}_0, \dots, \hat{\beta}_e)$ in table 5 varies from those of ℓ_1 -norm since the second one measures the main trend regression function and the third one measures only a part or a period of years observations.

5. Appendix

Proof of Proposition 2.2. Let $m_{pq} = \min \{p, q\}$. For $0 \leq q \leq \sum_{i=1}^p a_i - 1$, denote piecewise polynomial

$$f_q = f_q^1 + f_q^2, \tag{5.1}$$

where

$$\begin{aligned} f_q^1 &= p_0^{0\dots 0} + \sum_{s=1}^p \sum_{t_s}^{\min\{q, a_s\}} p_{t_s}^{0\dots t_s \dots 0} I(x_s > \delta_{t_s}^s) \\ &+ \sum_{1 \leq s_1 < s_2 \leq p} \sum_{\substack{t_{s_1} + t_{s_2} = 2 \\ 1 \leq t_{s_i} \leq a_{s_i}}}^q p_{t_{s_1} + t_{s_2}}^{0\dots t_{s_1} \dots t_{s_2} \dots 0} I(x_{s_1} > \delta_{t_{s_1}}^{s_1}, x_{s_2} > \delta_{t_{s_2}}^{s_2}) \\ &\vdots \\ &+ \sum_{1 \leq s_1 < \dots < s_{m_{pq}} \leq p} \sum_{\substack{t_{s_1} + \dots + t_{s_{m_{pq}}} = q \\ 1 \leq t_{s_i} \leq a_{s_i}}}^{0\dots t_{s_1} \dots t_{s_2} \dots t_{s_{m_{pq}}} \dots 0} p_{t_{s_1} + \dots + t_{s_{m_{pq}}}}^{0\dots t_{s_1} \dots t_{s_2} \dots t_{s_{m_{pq}}} \dots 0} \\ &I(x_{s_i} > \delta_{t_{s_i}}^{s_i}, i = 1, \dots, m_{pq}) \end{aligned}$$

and

$$f_q^2 = \sum_{t_1 + \dots + t_p = q+1}^{a_1 + \dots + a_p} p_q^{t_1 \dots t_p} I(\delta_{t_i}^i < x_i \leq \delta_{t_{i+1}}^i, i = 1, \dots, p),$$

where p_q are tensor product polynomials. We further let

$$f_q^{20} = \sum_{t_1 + \dots + t_p = q+1} p_q^{t_1 \dots t_p} I(\delta_{t_i}^i < x_i \leq \delta_{t_{i+1}}^i, i = 1, \dots, p).$$

Consider the decomposition $f_q^{20} = B_q + (f_q^{20} - B_q)$ where

$$\begin{aligned} B_q &= \sum_{s=1}^p p_q^{0 \dots q+1 \dots 0} I(x_s > \delta_{q+1}^s, q+1 \leq a_s) \\ &+ \sum_{1 \leq s_1 < s_2 \leq p} \sum_{\substack{t_{s_1} + t_{s_2} = q+1 \\ 1 \leq t_{s_i} \leq a_{s_i}}} p_q^{0 \dots t_{s_1} \dots t_{s_2} \dots 0} I(x_{s_i} > \delta_{t_{s_i}}^{s_i}, i = 1, 2) \\ &\vdots \\ &+ \sum_{1 \leq s_1 < \dots < s_{m_{pq}} \leq p} \sum_{\substack{t_{s_1} + \dots + t_{s_{m_{pq}}} = q+1 \\ 1 \leq t_{s_i} \leq a_{s_i}}} p_q^{0 \dots t_{s_1} \dots t_{s_{m_{pq}}} \dots 0} I(x_{s_i} > \delta_{t_{s_i}}^{s_i}, \\ &i = 1, \dots, m_{pq}). \end{aligned}$$

We will show that for each tensor product piecewise polynomial f and $0 \leq q \leq \sum_{i=1}^p a_i - 1$ there exists a function f_q of (5.1) such that $f = f_q$. Let $q = 0$. Obviously,

$$f = p^{0 \dots 0} + \sum_{t_1 + \dots + t_p = 1}^{a_1 + \dots + a_p} p_1^{t_1 \dots t_p} I(\delta_{t_i}^i < x_i \leq \delta_{t_{i+1}}^i, i = 1, \dots, p)$$

with $p_1^{t_1 \dots t_p} = p^{t_1 \dots t_p} - p^{0 \dots 0}$. Denote by $p_0^{0 \dots 0} = p^{0 \dots 0}$. Then $f = f_0$. Suppose that $f = f_q (= f_q^1 + f_q^2)$. Can check that $f_q^{20} - B_q$ and $f_q^2 - f_q^{20}$ are all tensor product piecewise polynomials defined on the region

$$\cup_{t_1 + \dots + t_p = q+2}^{a_1 + \dots + a_p} \{(x_1, \dots, x_p) : \delta_{t_i}^i < x_i \leq \delta_{t_{i+1}}^i, i = 1, \dots, p\}.$$

We now define $p_{q+1}^{t_1 \dots t_p}$. Denote polynomial $p_q^{t_1 \dots t_p}$ by $p_{q+1}^{t_1 \dots t_p}$ for which $p_q^{t_1 \dots t_p}$ is the polynomial of B_q defined on region $\{(x_1, \dots, x_p) : \delta_{t_i}^i < x_i, i = 1, \dots, p\}$. For (t_1, \dots, t_p) be such that $\sum_{i=1}^p t_i \geq q+2$, let $p_{q+1}^{t_1 \dots t_p}$ be such that $f_{q+1}^2 = f_q^2 - B_q$. We can also see that $f_q^1 + B_{q+1} = f_{q+1}^1$. Then we have

$$\begin{aligned} f &= f_q^1 + B_q + f_q^2 - B_q \\ &= f_{q+1}^1 + f_{q+1}^2 \\ &= f_{q+1}. \end{aligned}$$

The proof is done by setting $q = \sum_{i=1}^p a_i - 1$.

Proof of Theorem 3.1. We will prove only part (a) while the proof of (b) is similar. Denote by $\pi y_i = \prod_{i=1}^p y_i$ and $\tilde{y} = y_1, \dots, y_p$ for any finite set $\{y_1, \dots, y_p\}$. Define a set of polynomial functions and truncated polynomial functions ψ by

$$\psi_{\pi s_i, \pi c_i}(\tilde{x}) = (\pi c_i)^{-1} \prod_{i=1}^p [(x_i - \delta_{s_i}^i)^{c_i}_+ I(s_i > 0) + x_i^{c_i} I(s_i = 0)], \quad (5.2)$$

$$0 \leq c_i \leq k, s_i = 0, 1, \dots, a_i, i = 1, \dots, p.$$

The elements in (5.2) is also with number $(k + 1)^p \prod_{i=1}^p (a_i + 1)$ and each element of them can be formulated as a linear combination of functions in (1.2). So the proof of this theorem is finished if we can show that the functions ψ of (5.2) are linearly independent. We then finish the proof by showing that any zero linear combination of functions in (5.2) will have zero coefficients. For $t_i = -, +, i = 1, \dots, p$, let the partial derivative by

$$\psi^{\pi c_i^0}(\delta_{s_1^0}^{t_1}, \dots, \delta_{s_p^0}^{t_p}) = \frac{\partial^{\sum_{i=1}^p c_i^0}}{\pi \partial x_i^{c_i^0}} \psi_{\pi s_i, \pi c_i}(\tilde{x})|_{x_i = \delta_{s_i^0}^{t_i}, i = 1, \dots, p,$$

here “+” and “-” represent the “right” and “left” derivative, respectively.

We also define linear function $\lambda_{\pi s_i^0, \pi c_i^0}$ on ψ by

$$\lambda_{\pi s_i^0, \pi c_i^0}(\psi) = \sum_{t_p = -, +} \dots \sum_{t_1 = -, +} (-1)^{\sum_{i=1}^p \ell(t_i)} \psi^{\pi c_i^0}(\delta_{s_1^0}^{t_1}, \dots, \delta_{s_p^0}^{t_p}), \quad (5.3)$$

where ℓ is the binary function defined by $\ell(t) = 1$ if $t = +$ and 0 if $t = -$.

If we let

$$h_m(x_m) = \sum_{\substack{t_i = -, + \\ i = 1, \dots, p \\ i \neq m}} (-1)^{\sum_{i \neq m} \ell(t_i)} \psi^{\pi c_i^0}(\delta_{s_1^0}^{t_1}, \dots, x_m, \dots, \delta_{s_p^0}^{t_p})$$

has continuous c_m th derivative at $\delta_{s_m^0}$, then

$$\lambda_{\pi s_i^0, \pi c_i^0}(\psi) = -\frac{\partial^{c_m}}{\partial x_m^{c_m}} h_m(x_m)|_{x_m = \delta_{s_m^0}^+} + \frac{\partial^{c_m}}{\partial x_m^{c_m}} h_m(x_m)|_{x_m = \delta_{s_m^0}^-} = 0.$$

With careful inspection, it can be seen that

$$\lambda_{\pi s_i^0, \pi c_i^0}(\psi) = \begin{cases} 1 & \text{if } s_i^0 = s_i, c_i^0 = c_i, i = 1, \dots, p \\ 0 & \text{otherwise} \end{cases}$$

This shows that the set $\{\psi_{\pi s_i, \pi c_i}(\bar{x})\}$ is linearly independent and then the set of functions in (1.2) is a basis of the space of order k tensor product piecewise polynomials.

Proof of Theorem 3.2. We only prove the case with $+$ function. The functions in (3.1) with $*$ = $+$ and c_i, c_s, c_{s_1} and $c_{s_2} = j + 1, \dots, k$ is a subset of the basis of the space of tensor product piecewise polynomials. Denote this set by D_j . We know that D_j is a linearly independent set. It is also obvious that each element in D_j is continuous for its j -th partial derivative. We then need only to show that these functions in D_j forms a generator of smoothness- j tensor product polynomial splines. The proof follows the proof of the linear independence of the set of (3.1) in Theorem 3.1.

References

Breiman, L. (1991), Discussion on Multivariate Adaptive Regression Splines, by Friedman, *The Annals of Statistics*, 19, 82-91.

Buse, A. and Lim, L. (1977), Cubic Splines as a Special Case of Restricted Least Squares, *Journal of the American Statistical Association*, 72, 64-68.

Chen, L. A. (1996), Bivariate Regression Splines, *Computational Statistics and Data Analysis*, 21, 399-418.

Chen, L. A. (1996), Symmetric Type Two-Stage Trimmed Mean for Simultaneous Equations Model, Technical report.

Chen, L. A. and Portnoy, S. (1996), Two-Stage Regression Quantiles and Two-Stage Trimmed Least Squares Estimators for Simultaneous Equations Models, *Communication in Statistics: Theory and Method*, 25, 1005-1032.

Cleveland, W. S. and Devlin, S. J. (1988), Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting, *Journal of the American Statistical Association*, 83, 596-610.

- Cleveland, W. S. and Mallows, C. L. and McRae, J. E. (1993), ATS Methods: Nonparametric Regression for Non-Gaussian Data, *Journal of the American Statistical Association*, 88, 821-835.
- De Boor, C. (1978), *A Practical Guide to Splines*, New York: Springer-Verlag.
- Eubank, R. L. (1988), *Spline Smoothing and Nonparametric Regression*, New York: Marcel Dekker, Inc.
- Friedman, J. H. (1991), Multivariate Adaptive Regression Splines, *The Annals of Statistics*, 19, 1-67.
- Friedman, J. H. and Silverman, B. W. (1989), Flexible Parsimonious Smoothing and Additive Modeling, (with discussion), *Technometrics*, 31, 3-39.
- Gu, C. and Wahba, G. (1993), Semiparametric Analysis of Variance with Tensor Product Thin Plate Splines, *Journal of the Royal Statistical Society*, B. 55, 353-368.
- Hastie, T., Tibshirani, R. and Andreas, B. (1994), Flexible Discriminant Analysis by Optimal Scoring, *Journal of the American Statistical Association*, 89, 1255-1270.
- Koenker, R. and d'Orey, V. (1987), Computing Regression Quantiles, *Applied Statistics*, 36, 383-393.
- LeBlanc, M. and Tibshirani, R. (1994), Adaptive Principal Surfaces, *Journal of the American Statistical Association*, 89, 53-64.
- Maddala, G. S. (1988), *Introduction to Econometrics*, New York: Macmillan Publishing Company.
- Poirier, D. J. (1973), Piecewise Regression Using Cubic Splines, *Journal of the American Statistical Association*, 68, 515-524.
- Smith, P. L. (1979), Splines as a Useful and Convenient Statistical Tool, *The American Statistician*, 33, 57-62.
- Stone, C. J. (1994), The Use of Polynomial Splines and Their Tensor Products in Multivariate Function Estimation, (with discussions), *The Annals of Statistics*, 22, 118-184.

- Theil, H. (1971), *Principles of econometrics*, New York: Wiley.
- Wahba, G. (1979), Convergence Rates of Thin Plate Smoothing Splines When the Data are Noisy, in *Smoothing Techniques for Curve Estimation*, eds, T. Gasser and M. Rosenblatt, (Springer-Verlag, Berlin) 233-245.
- Wegman, E. J. and Wright, I. W. (1983), Spline in Statistics, *Journal of the American Statistical Association*, 78, 351-365.
- Zhang, H. (1994), Maximal Correlation and Adaptive Splines, *Technometrics*, 36, 196-201.

Received April, 1997.