KATHOLIEKE
UNIVERSITEIT
LEUVEN

DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

RESEARCH REPORT 0118

NONPARAMETRIC TIME TRENDS IN OPTIMAL DESIGN OF EXPERIMENTS

by
L. TACK
M. VANDEBROEK

D/2001/2376/18

Nonparametric Time Trends in Optimal Design of Experiments

L. Tack
M. Vandebroek
Katholieke Universiteit Leuven, Belgium

When performing an experiment, the observed responses are often influenced by a temporal trend due to aging of material, learning effects, equipment wear-out, warm-up effects, etc. The construction of run orders that are optimally balanced for time trend effects relies on the incorporation of a parametric representation of the time dependence in the response model. The parameters of the time trend are then treated as nuisance parameters. However, the price one has to pay for by purely parametric modeling is the biased results when the time trend is misspecified. This paper presents a design algorithm for the construction of optimal run orders when kernel smoothing is used to model the temporal trend nonparametrically. The benefits of modeling the time trend nonparametrically are outlined. Besides, the influence of the bandwidth and the kernel function on the performance of the optimal run orders is investigated. The presented design algorithm shows to be very useful when it is hard to model the time dependence parametrically or when the functional form of the time trend is unknown. An industrial example illustrates the practical utility of the proposed design algorithm.

Keywords: $\mathcal{D}_t$-optimality; exchange algorithm; kernel smoothing; run order; nonparametric regression; simulated annealing; time trend;

1 Introduction

Performing experiments in a time sequence often creates time order dependence in the observed responses. Incorporation of time trend effects results into the mathematical model

$$y = \mathbf{f}'(\mathbf{x})\boldsymbol{\alpha} + \mathbf{g}'(t)\boldsymbol{\beta} + \varepsilon, \tag{1}$$

with $\mathbf{f}(\mathbf{x})$ the $p \times 1$ vector representing the polynomial expansion of $\mathbf{x}$ for the response model, $\mathbf{g}(t)$ the $q \times 1$ vector representing the polynomial expansion for the time trend, expressed as a function of time $t \in [-1, 1]$, $\boldsymbol{\alpha}$ the $p \times 1$ vector of important parameters and $\boldsymbol{\beta}$ the $q \times 1$ vector of parameters of the polynomial time trend. The independent error terms ε are assumed to have expectation zero and constant variance σ^2. In this paper, no interaction effects between $\mathbf{x}$ and t are considered, an assumption which holds in many practical situations. For n observations, it is convenient to rewrite (1) as

$$\mathbf{y} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{G}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

with $\mathbf{y}$ an n-dimensional vector of observed responses and $\mathbf{F}$ and $\mathbf{G}$ the $n \times p$ and the $n \times q$ design matrices respectively. When primary interest is in the precision of the parameter estimates, the construction of run orders that are optimally balanced for time trends is based on maximization of the information on the important parameters $\boldsymbol{\alpha}$, whereas the parameters $\boldsymbol{\beta}$ modeling the time trend are treated as nuisance parameters. The resulting run order is called the $\mathcal{D}_t$-optimal run order $\delta_{\mathcal{D}_t}$. The value of the optimality criterion equals

$$\mathcal{D}_t = \frac{\begin{vmatrix} \mathbf{F}'\mathbf{F} & \mathbf{F}'\mathbf{G} \\ \mathbf{G}'\mathbf{F} & \mathbf{G}'\mathbf{G} \end{vmatrix}}{|\mathbf{G}'\mathbf{G}|} = |\mathbf{F}'\mathbf{F} - \mathbf{F}'\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{F}|. \tag{3}$$

In the absence of time trend effects, the $\mathcal{D}$-optimal design $\delta_{\mathcal{D}}$ maximizes the determinant of the information matrix, i.e. $\mathcal{D} = |\mathbf{F}'\mathbf{F}|$. Atkinson and Donev (1996) compare the $\mathcal{D}$-optimal and the $\mathcal{D}_t$-optimal design through the trend factor

$$\mathrm{TF}(\delta_{\mathcal{D}_t}) = \left\{ \frac{\mathcal{D}_t(\delta_{\mathcal{D}_t})}{\mathcal{D}(\delta_{\mathcal{D}})} \right\}^{1/p}. \tag{4}$$

The power $1/p$ ensures that the trend factor has the dimension of variance. This means for instance that a design $\delta_{\mathcal{D}_t}$ with trend factor 0.5 has to be replicated twice in order to be equally informative as the $\mathcal{D}$-optimal design. Bradley and Yeh (1980) define a design to be trend-free if the treatment effects are orthogonal to the polynomial trend components. In the context of regression designs, this condition comes down to $\mathbf{F}'\mathbf{G} = \mathbf{0}$. As a result, a trend-free design has the maximum value of the trend factor, namely $\mathrm{TF}(\delta_{\mathcal{D}_t}) = 1$. In situations where it is impossible to obtain completely trend-free run orders, $\mathrm{TF}(\delta_{\mathcal{D}_t})$ will be less than 1. Based on Atkinson and Donev (1996), Tack and

Vandebroek (2001) present a design algorithm for the construction of cost-efficient run orders that are optimally balanced for known time trends. For example, they compute optimal run orders for design problems in which drift deteriorates the sensitivity of a spectrophotometer or the performance of a polisher used in the electronics industry.

The next section illustrates that if the time dependence is misspecified, the computed $\mathcal{D}_t$-optimal run orders become less useful.

2 Misspecification of the time trend

Every good book on econometric models shows that excluding relevant variables from a statistical model leads to biased and inconsistent parameter estimates. On the other hand, including an irrelevant variable in the model gives unbiased and consistent parameter estimates but the estimated variances of the parameters are larger. In this paper, we assume that the response model $\mathbf{f}$ in (1) is specified correctly and we confine ourselves to potential misspecification of the time trend. The next two sections describe the effects of misspecification of the time trend on the bias and the precision of the important parameter estimates $\hat{\boldsymbol{\alpha}}$.

2.1 Bias of important parameter estimates

Consider as an example an experimenter who postulates a polynomial time trend of order $q = q_1$. The assumed statistical model then equals

$$\mathbf{y} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{G}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}, \tag{5}$$

with $\mathbf{G}_1$ the $n \times q_1$ extended design matrix for the time trend and $\boldsymbol{\beta}_1$ the vector of q_1 trend parameters. However, if the true time trend contains $q_1 + q_2$ parameters, the true regression model is

$$\mathbf{y} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{G}_1\boldsymbol{\beta}_1 + \mathbf{G}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \tag{6}$$

with $\mathbf{G}_2$ and $\boldsymbol{\beta}_2$ related to the q_2 trend parameters discarded by the experimenter. Based on (5) and (6), it is proven in appendix 1 that the bias introduced by misspecification of the time trend is equal to

$$\mathrm{E}(\hat{\boldsymbol{\alpha}}) - \boldsymbol{\alpha} = [\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{F}]^{-1}\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{G}_2\boldsymbol{\beta}_2. \tag{7}$$

This means that underspecifying the order of the time trend leads to biased parameter estimates $\hat{\boldsymbol{\alpha}}$. For a trend-free design, $\mathbf{F}'\mathbf{G}_1 = \mathbf{0}$ and (7) simplifies to

$$\mathrm{E}(\hat{\boldsymbol{\alpha}}) - \boldsymbol{\alpha} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{G}_2\boldsymbol{\beta}_2. \tag{8}$$

As a matter of fact, if the important effects would also be orthogonal to the q_2 omitted trend parameters, the bias (8) disappears.

On the other hand, if the fitted response model is

$$\mathbf{y} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{G}_1\boldsymbol{\beta}_1 + \mathbf{G}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \tag{9}$$

and the true regression model

$$\mathbf{y} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{G}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}, \tag{10}$$

the parameter estimates $\hat{\boldsymbol{\alpha}}$ are unbiased. The proof is given in appendix 2. However, including irrelevant parameters leads to inefficient estimates. The next section will illustrate the loss in precision of the parameter estimates if the order of the trend is overspecified.

2.2 Precision of important parameter estimates

Besides biased parameter estimates resulting from model underspecification, the following example illustrates how overspecification of the temporal trend leads to a decrease in the precision of the important parameter estimates $\hat{\boldsymbol{\alpha}}$. An experiment is conducted to evaluate the homogeneity of the amount of nitrogen in steel rods. The steel rods are produced on a blast-furnace and the three variables under study are the reheating temperature in degrees Celsius, the amount of carbon added in terms of percentage and the amount of chromium added in terms of percentage. The number of observations equals 20 and the coded factor levels are shown in table 1.

Table 1: Factors and factor levels in the nitrogen experiment

factor	coded levels
reheating temperature x_1	-1, 1
amount of carbon x_2	-1, -0.78, 1
amount of chromium x_3	-1, 0.4, 1

The response model is represented by $\mathbf{f}'(\mathbf{x}) = [1, x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_2^2, x_3^2]$ and the observations are expected to be influenced by time trend effects because previous work revealed that the nitrogen contents decreased drastically as the day went on. $\mathcal{D}_t$-optimal run orders will be constructed for the following postulated time trend models:

$$\begin{aligned} \mathbf{g}_1'(t) &= [t], \\ \mathbf{g}_2'(t) &= [t, t^2], \\ \mathbf{g}_3'(t) &= [t, t^2, t^3]. \end{aligned}$$

As an illustration the $\mathcal{D}_t$-optimal run orders are shown in table 2. The computed $\mathcal{D}_t$-optimal run order for time trend $\mathbf{g}_1$ is completely trend-free. However, the trend factor of the $\mathcal{D}_t$-optimal run order for time trend $\mathbf{g}_2$ equals 0.913 and that of the $\mathcal{D}_t$-optimal run order for time trend $\mathbf{g}_3$ equals 0.905. Consequently, one observes that the $\mathcal{D}_t$-value of the

Table 2: $\mathcal{D}_t$-optimal run orders for different time trends in the nitrogen experiment

	$\mathbf{g}_1(t)$			$\mathbf{g}_2(t)$			$\mathbf{g}_3(t)$		
run	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
1	1	1	0.4	-1	-0.78	-1	-1	-1	0.4
2	-1	-1	1	-1	1	1	1	-0.78	-1
3	-1	1	-1	1	-1	0.4	-1	1	-1
4	1	-0.78	0.4	1	1	-1	1	1	1
5	-1	-1	-1	1	-1	1	1	-0.78	1
6	-1	-0.78	1	1	1	-1	-1	-1	1
7	1	1	-1	-1	-0.78	0.4	-1	1	0.4
8	1	1	1	-1	1	1	1	1	-1
9	1	-0.78	1	-1	-1	0.4	-1	-0.78	0.4
10	-1	1	0.4	-1	1	-1	1	-1	-1
11	-1	-0.78	1	1	1	1	-1	-1	-1
12	-1	1	1	-1	-0.78	1	1	-0.78	0.4
13	1	-1	1	-1	-1	-1	-1	1	-1
14	1	-0.78	-1	1	-1	-1	1	1	0.4
15	-1	-0.78	-1	1	-0.78	1	1	-1	1
16	-1	1	-1	-1	1	0.4	-1	-0.78	1
17	1	-1	0.4	1	-0.78	-1	-1	1	1
18	1	1	-1	-1	1	-1	1	1	-1
19	1	1	1	1	1	0.4	-1	-0.78	-1
20	-1	-1	0.4	-1	-1	1	1	-1	0.4

optimal run orders diminishes as the order of the postulated time trend becomes larger. Alternatively stated, the larger the order of the time trend, the more information is lost on the important parameters.

It is also important to investigate how the computed run orders perform when the true temporal trend differs from the postulated one. For the sake of clarity, the true time trend will be denoted as $\tilde{\mathbf{g}}$. For several polynomial time trends, table 3 shows the difference in terms of percentage between the trend factor of the $\mathcal{D}_t$-optimal run order when the time trend is specified correctly as $\tilde{\mathbf{g}}$ and the trend factor of the run order computed with postulated time trend $\mathbf{g}_i$. This difference in trend factor is used as a measure of the difference in precision of the important parameter estimates. For instance, misspecification of a true quadratic time trend $\tilde{\mathbf{g}} = \mathbf{g}_2$ as a linear time trend $\mathbf{g}_1$ involves a decrease in the trend factor of about 8.13%. Remark that the decrease in information due to misspecification of the order of the time trend especially comes true when the true time trend $\tilde{\mathbf{g}}$ is of higher order than the postulated one. Consequently, the nitrogen experiment shows that due to misspecification of the order of the time trend one can end up with a $\mathcal{D}_t$-optimal run order that is far from the $\mathcal{D}_t$-optimal run order for the true time trend.

Table 3: Relative difference in trend factor between $\mathcal{D}_t$-optimal run orders for correct specification of the time trend and misspecification of the order of the time trend

postulated time trend	true time trend $\tilde{\mathbf{g}} = \mathbf{g}_1$	$\tilde{\mathbf{g}} = \mathbf{g}_2$	$\tilde{\mathbf{g}} = \mathbf{g}_3$
$\mathbf{g}_1$	0	8.13	8.77
$\mathbf{g}_2$	0.01	0	11.01
$\mathbf{g}_3$	0.52	0.73	0

As a conclusion, the bias results of section 2.1 and the precision results of section 2.2 clearly illustrate the need to develop design methodologies that are less dependent on the assumed functional form of the model and that better guard against misspecification of the time trend. Besides, many time trend effects due to aging or wear-out have a complicated behaviour and the experimenter would benefit from design strategies that no longer need a specification of the functional form of the time trend. In this paper, the main emphasis is on the construction of run orders that are optimally balanced for nonparametric time trends.

A detailed description of nonparametric and semiparametric regression techniques can be found in Härdle (1990). In this paper we confine ourselves to the widely spread and well known kernel smoothing technique used in nonparametric regression. The next section shortly reviews kernel regression and section 4 gives a literature review of the application of nonparametric regression techniques in optimal design. Section 5 introduces a new design criterion to compute optimal run orders in the presence of nonparametric time trends and our design algorithm is outlined in section 6. Section 7 illustrates the benefits that result from modeling the time trend nonparametrically. Finally, section 8 provides a practical example.

3 Kernel smoothing in nonparametric regression

Nonparametric regression forms a collection of techniques for estimating a regression curve without making strong assumptions about the shape of the true regression function. Suppose we are given n observations $\{(x_i, y_i)\}_{i=1}^n$ satisfying the model

$$y_i = f(x_i) + \varepsilon_i, \tag{11}$$

where $\{\varepsilon_i\}_{i=1}^n$ are independent random error terms with zero mean and constant variance σ^2 and f is an unknown function. Assume that $a \leq x_1 < \cdots < x_n \leq b$ for finite constants a and b. The aim of nonparametric regression is to produce a reasonable estimate of the unknown function f. This curve estimation procedure is commonly called smoothing and the idea is based on local averaging. This means that an estimate of y at point x is given

by

$$\hat{y}(x) = \sum_{i=1}^{n} w_i(x) y_i, \tag{12}$$

where $\{w_i(x)\}_{i=1}^n$ denotes a sequence of weights that depend on $\{x_i\}_{i=1}^n$. Generally speaking, the further away the points x_i from x, the smaller the weight $w_i(x)$. This means that only points x_i in the neighbourhood of x are given a non-zero weight $w_i(x)$. Because the estimate $\hat{y}$ is a linear combination of all y_i, these smoothers are often referred to as linear smoothers. Based on (12), the n-dimensional vector of estimates $\hat{y}(x_i)$ can be written as

$$\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}, \tag{13}$$

where $\mathbf{S}$ is called the smoother matrix with elements $s_{ij} = w_j(x_i)$ and $i, j \in \{1, \ldots, n\}$. The amount of averaging is controlled by the weight sequence $\{w_i(x)\}_{i=1}^n$ which is tuned by a smoothing parameter. This smoothing parameter regulates the size of the neighbourhood around x and balances the degree of fidelity to the data against the smoothness of the estimated curve.

Among several methods for choosing the weights $w_i(x)$, we confine ourselves to kernel smoothing because it is one of the simplest ways to compute the weight sequence. In kernel regression, the smoothing parameter is referred to as the bandwidth λ and the weights are defined as

$$w_i(x) = \frac{K\left(\frac{x_i - x}{\lambda}\right)}{\sum_{j=1}^{n} K\left(\frac{x_j - x}{\lambda}\right)}, \tag{14}$$

with K the kernel function satisfying

$$\int_{-1}^{1} K(u)du = 1,$$

$$\int_{-1}^{1} x^l K(u)du = 0,$$

for $l = 1, \ldots, m-1$ and m called the order of the kernel. The support of K is on [-1,1]. The resulting estimator (12) is often referred to as the Nadaraya-Watson estimator and the denominator of (14) divided by $n\lambda$ is called the Rosenblatt-Parzen kernel density estimator.

Generally speaking, an increased bandwidth λ results into an oversmooth curve with decreased variance of $\hat{y}$, but with increased bias. On the other hand, defining the smoothing parameter so that it corresponds to a very small neighbourhood would make $\hat{y}$ very wiggly. In this case the variability of $\hat{y}$ would be inflated. The trade-off between bias and variance can be made more precise by investigating the large sample properties of the

regression estimators. If the sample size n goes to infinity, the bandwidth λ to zero and $n\lambda$ to infinity, it can be shown that the mean squared error $\mathrm{E}\{\hat{y}(x) - y(x)\}^2$ is equal to

$$\lambda^{2m}\left\{f^{(m)}(x)\int_{-1}^{1} u^m K(u)du/m!\right\}^2 + o(\lambda^{2m}) + \sigma^2\int_{-1}^{1} K^2(u)du/n\lambda + o(1/n\lambda). \tag{15}$$

The first two terms denote the squared bias $[\mathrm{E}\{\hat{y}(x)\} - y(x)]^2$ and the last two terms represent the variance of $\hat{y}(x)$. One easily sees that the bias disappears when the bandwidth λ goes to zero and the variance disappears when $n\lambda$ goes to infinity. Gasser *et al.* (1985) derive expressions for kernels K that minimize the asymptotic variance. They also derive expressions for optimal kernels K, i.e. kernels that minimize the optimal integrated mean squared error $\int_a^b \mathrm{E}\{\hat{y}(x) - y(x)\}^2 dx$. The optimal integrated mean squared error is calculated as the integrated mean squared error with a bandwidth λ for which the integrated mean squared error is a minimum. Table 4 provides examples of minimum variance kernels and optimal kernels. Note that kernel function K_4 is often referred to as the Epanechnikov kernel (Epanechnikov, 1969).

Table 4: Examples of minimum variance kernels and optimal kernels

	order m	minimum variance kernel
K_1	2	$1/2$
K_2	4	$3/8(-5u^2+3)$
K_3	6	$15/128(63u^4-70u^2+15)$
	order m	optimal kernel
K_4	2	$3/4(-u^2+1)$
K_5	4	$15/32(7u^4-10u^2+3)$
K_6	6	$35/256(-99u^6+189u^4-105u^2+15)$

The accuracy of the estimated curve is not only a function of the kernel K, but depends also on the bandwidth λ. Various data-driven methods such as cross-validation (Clark, 1977) and generalized cross-validation have been proposed for choosing the bandwidth parameter λ that minimizes a function of the mean squared error. Ordinary cross-validation leaves the data points out one at a time and chooses that value of λ under which the missing data point is best predicted by the remainder of the data. Craven and Wahba (1979) introduced generalized cross-validation in which the optimal bandwidth λ is chosen so as to minimize

$$\mathrm{GCV}(\lambda) = \frac{n\sum_{i=1}^{n}\{y_i - \hat{y}_i(\mathbf{x}_i)\}^2}{\{n - \mathrm{trace}(\mathbf{H})\}^2}, \tag{16}$$

with $\mathbf{H}$ the hat matrix defined by $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$. In nonparametric regression, the hat matrix $\mathbf{H}$ equals the smoother matrix $\mathbf{S}$. Craven and Wahba (1979) show that, when n goes

to infinity, the mean squared error with λ determined from minimizing (16), tends to a minimum. Calculating the optimal bandwidth based on generalized cross-validation is computationally less expensive than ordinary cross-validation and Rice (1984) applies generalized cross-validation in kernel smoothing. Because of the fact that prior to the experiments no observations y_i are available, these data-driven methods cannot be used to select an appropriate bandwidth when designing an experiment. This problem will be dealt with in section 8. The next section gives a literature overview of nonparametric design of experiments.

4 Nonparametric optimal design

A few approaches to nonparametric optimal design have been encountered in the literature. Müller (1984) uses the optimal integrated mean squared error as the criterion to be minimized and shows that for univariate and fixed kernel regression, i.e. where the same bandwidth is used for any estimation point x, the optimal density of the design points is uniform. Müller (1984) also shows that when the bandwidth is a function of x, then asymptotically the optimal design density is some function of the roughness of the unknown function f and of the order of the kernel function used.

Based on Müller (1984), Faraway and Rothman (1989) tackle the problem of sequentially selecting the design points for a nonparametric regression problem. In sequential design, the decision on the position of the next design point is based on the previous observations. The advantage of sequential design is that significantly larger precision can be obtained for the same sample size as used in non-sequential experimental design. Faraway and Rothman (1989) consider univariate kernel regression by which the design points are chosen so as to make the density of the actual design points as close as possible to the estimated optimal density of the design points. They present a method that both selects the local bandwidth required for the regression estimate and the optimal location of the next best design point. Faraway (1990) extends this method to the nonparametric estimation of surfaces.

Another application of univariate nonparametric regression in optimal design of experiments comes from Butler (1989). He shows that $\mathcal{G}$-optimal polynomial designs are also $\mathcal{G}$-optimal for smoothing splines[1] if the smoothing parameter is sufficiently large. For smaller values of the smoothing parameter, in a number of cases the optimal designs can be calculated analytically. Butler (1990) derives multivariate optimal densities of the design points for the weighted $\mathcal{V}$-optimality criterion.

The next section elaborates our approach to the nonparametric representation of the time

[1] In spline smoothing, the best compromise between goodness of fit and roughness is obtained by minimizing $\sum_{i=1}^{n}\{y_i - f(x_i)\}^2 + \lambda \int_a^b \{f^{(2)}(u)\}^2 du$. The cubic spline is the unique solution to this minimization problem and is composed of cubic polynomials between any two successive values x_i and x_{i+1}. Besides, the cubic spline and its first two derivatives are continuous at the observation points and at the boundary points x_1 and x_n the second derivative of $\hat{f}$ is zero.

dependence. Because in this paper primary interest lies in the precision with which the parameters $\boldsymbol{\alpha}$ are estimated, we present an extension of the $\mathcal{D}_t$-optimality criterion to semiparametric regression designs.

5 Semiparametric optimal design

For the parametric model (2), the normal equations for determining $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ equal

$$\mathbf{F}'\mathbf{F}\boldsymbol{\alpha} = \mathbf{F}'(\mathbf{y} - \mathbf{G}\boldsymbol{\beta}) \tag{17}$$

and

$$\mathbf{G}'\mathbf{G}\boldsymbol{\beta} = \mathbf{G}'(\mathbf{y} - \mathbf{F}\boldsymbol{\alpha}). \tag{18}$$

In view of what will follow, we rewrite (18) as

$$\mathbf{G}\boldsymbol{\beta} = \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'(\mathbf{y} - \mathbf{F}\boldsymbol{\alpha}). \tag{19}$$

However, when no assumptions are made about the functional shape of the time trend, the model for the observed responses equals

$$y = \mathbf{f}'(\mathbf{x})\boldsymbol{\alpha} + g(t) + \varepsilon, \tag{20}$$

with $g(t)$ an unknown function representing the time trend. For n observations, the partially linear or semiparametric model (20) can be rewritten as

$$\mathbf{y} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{g} + \boldsymbol{\varepsilon},$$

with time points $t_1, \ldots, t_n$ and $\mathbf{g}' = [g(t_1), \ldots, g(t_n)]$. In this case, there is no parameter vector $\boldsymbol{\beta}$ and the normal equations (17) and (19) become meaningless. Therefore, Green *et al.* (1985) propose to adapt (17) and (19) to

$$\mathbf{F}'\mathbf{F}\boldsymbol{\alpha} = \mathbf{F}'(\mathbf{y} - \mathbf{g}) \tag{21}$$

and

$$\mathbf{g} = \mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'(\mathbf{y} - \mathbf{F}\boldsymbol{\alpha}) \tag{22}$$

respectively. Besides, they replace the projection matrix $\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$ by the smoother matrix $\mathbf{S}$ defined in (13) and obtain

$$\mathbf{g} = \mathbf{S}(\mathbf{y} - \mathbf{F}\boldsymbol{\alpha}). \tag{23}$$

For kernel regression, the smoother matrix $\mathbf{S}$ is given by

$$\mathbf{S} = \begin{bmatrix} \frac{K\left(\frac{t_1 - t_1}{\lambda}\right)}{\sum_{i=1}^{n} K\left(\frac{t_i - t_1}{\lambda}\right)} & \cdots & \frac{K\left(\frac{t_n - t_1}{\lambda}\right)}{\sum_{i=1}^{n} K\left(\frac{t_i - t_1}{\lambda}\right)} \\ \vdots & \vdots & \vdots \\ \frac{K\left(\frac{t_1 - t_n}{\lambda}\right)}{\sum_{i=1}^{n} K\left(\frac{t_i - t_n}{\lambda}\right)} & \cdots & \frac{K\left(\frac{t_n - t_n}{\lambda}\right)}{\sum_{i=1}^{n} K\left(\frac{t_i - t_n}{\lambda}\right)} \end{bmatrix}. \tag{24}$$

The Green-Jennison-Seheult estimators (Green *et al.*, 1985) are then

$$\hat{\boldsymbol{\alpha}}_{GJS} = (\mathbf{F}'(\mathbf{I}-\mathbf{S})\mathbf{F})^{-1}\mathbf{F}'(\mathbf{I}-\mathbf{S})\mathbf{y}, \tag{25}$$

$$\hat{\boldsymbol{g}}_{GJS} = \mathbf{S}(\mathbf{y}-\mathbf{F}\hat{\boldsymbol{\alpha}}_{GJS}), \tag{26}$$

and the variance-covariance matrix of $\hat{\boldsymbol{\alpha}}_{GJS}$ equals

$$\text{COV}(\hat{\boldsymbol{\alpha}}_{GJS}) = \sigma^2(\mathbf{F}'\tilde{\mathbf{F}})^{-1}\mathbf{F}'(\mathbf{I}-\mathbf{S})(\mathbf{I}-\mathbf{S})'\mathbf{F}(\mathbf{F}'\tilde{\mathbf{F}})^{-1}, \tag{27}$$

with $\tilde{\mathbf{F}}$ defined as $(\mathbf{I}-\mathbf{S})\mathbf{F}$. In this paper, run orders will be preferred that maximize the amount of information on the important parameters $\hat{\boldsymbol{\alpha}}_{GJS}$. Based on (27), the corresponding optimality criterion equals

$$\mathcal{D}_t = \frac{|\mathbf{F}'\tilde{\mathbf{F}}|^2}{|\mathbf{F}'(\mathbf{I}-\mathbf{S})(\mathbf{I}-\mathbf{S})'\mathbf{F}|}. \tag{28}$$

Speckman (1988) provides another way of estimating the parameters in model (20) by first adjusting $\mathbf{F}$ and $\mathbf{y}$ for t by defining $\tilde{\mathbf{y}} = (\mathbf{I}-\mathbf{S})\mathbf{y}$ and $\tilde{\mathbf{F}} = (\mathbf{I}-\mathbf{S})\mathbf{F}$ and then regressing the residual $\tilde{\mathbf{y}}$ on the residual $\tilde{\mathbf{F}}$. The resulting estimators are

$$\hat{\boldsymbol{\alpha}}_{pr} = (\tilde{\mathbf{F}}'\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}'\tilde{\mathbf{y}}, \tag{29}$$

$$\hat{\boldsymbol{g}}_{pr} = \mathbf{S}(\mathbf{y}-\mathbf{F}\hat{\boldsymbol{\alpha}}_{pr}). \tag{30}$$

The index **pr** refers to the fact that the estimate $\hat{\boldsymbol{\alpha}}_{pr}$ is computed by regression on partial residuals. The variance-covariance matrix of the parameters $\hat{\boldsymbol{\alpha}}_{pr}$ equals

$$\text{COV}(\hat{\boldsymbol{\alpha}}_{pr}) = \sigma^2(\tilde{\mathbf{F}}'\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}'(\mathbf{I}-\mathbf{S})(\mathbf{I}-\mathbf{S})'\tilde{\mathbf{F}}(\tilde{\mathbf{F}}'\tilde{\mathbf{F}})^{-1}. \tag{31}$$

In case of partial residuals the criterion to be maximized then equals

$$\mathcal{D}_t = \frac{|\tilde{\mathbf{F}}'\tilde{\mathbf{F}}|^2}{|\tilde{\mathbf{F}}'(\mathbf{I}-\mathbf{S})(\mathbf{I}-\mathbf{S})'\tilde{\mathbf{F}}|}. \tag{32}$$

It can easily be shown that if $\mathbf{S}$ is a symmetric idempotent matrix, i.e. $\mathbf{S} = \mathbf{S}' = \mathbf{S}^2$, then $\hat{\boldsymbol{\alpha}}_{GJS}$ and $\hat{\boldsymbol{\alpha}}_{pr}$ will be identical estimators. If the smoother matrix vanishes from (28) and (32), then both criteria come down to the $\mathcal{D}$-optimality criterion $|\mathbf{F}'\mathbf{F}|$. On the other hand, when the temporal trend is modeled parametrically, the smoother matrix $\mathbf{S}$ has to be replaced by $\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$ and both optimality criteria then equal the $\mathcal{D}_t$-optimality criterion (3). Finally, when the bandwidth λ goes to infinity, it can be verified that both criteria (28) and (32) simplify to $\mathcal{D}_t = |\mathbf{F}'(\mathbf{I}-\mathbf{S})\mathbf{F}|$ with $\mathbf{S} = (1/n)\mathbf{1}\mathbf{1}'$. However, a major drawback of both estimation procedures is that the resulting estimators are biased. Speckman (1988) theoretically proves that for the optimal bandwidth, the asymptotic bias of $\hat{\boldsymbol{\alpha}}_{pr}$ is of lower order than the bias of $\hat{\boldsymbol{\alpha}}_{GJS}$. Section 7 extensively investigates bias problems in semiparametric optimal design with finite sample sizes.

Remark that with the Green-Jennison-Seheult method and the method based on partial residuals the intercept term in response model $\mathbf{f}(\mathbf{x})$ is not estimable. Because the row elements of the smoother matrix (24) add up to one, the column corresponding to the intercept in $\tilde{\mathbf{F}} = (\mathbf{I}-\mathbf{S})\mathbf{F}$ would only contain zero entries. This inevitably leads to singular matrices $\mathbf{F}'\tilde{\mathbf{F}}$ and $\tilde{\mathbf{F}}'\tilde{\mathbf{F}}$ in (27) and (31) respectively. For that reason, in the sequel of this paper the intercept is omitted from the response model $\mathbf{f}(\mathbf{x})$.

The next section describes our proposed design algorithm for the construction of $\mathcal{D}_t$-optimal run orders in the presence of nonparametric time trend effects.

6 The design algorithm

Based on the $\mathcal{D}_t$-optimality criterion defined in (3), Tack and Vandebroek (2000a, 2000b, 2001) developed exchange algorithms for the construction of cost-efficient or budget constrained run orders that are optimally balanced for parametric time trends. In this section, we present a simulated annealing algorithm for the construction of $\mathcal{D}_t$-optimal run orders when the time trend is modeled nonparametrically.

6.1 Outline of the algorithm

The aim of the proposed exchange algorithm is the construction of optimal run orders by allocating a user-specified number of observations n selected from a candidate set of d distinct design points to n out of h available time points so as to maximize the value of the $\mathcal{D}_t$-optimality criterion (28) or (32). The allocation occurs with replacement since the number of design points d is usually smaller than the number of observations n. After reading the input, a starting run order is constructed by allocating n randomly chosen design points to n randomly chosen time points. Next, the starting run order is subject to iterative improvements by evaluating the effect on the criterion value of three different design changes. The first design change consists of the deletion of a randomly selected design point $\mathbf{x}_i$ at time point t_k and the addition of a design point $\mathbf{x}_j \neq \mathbf{x}_i$ randomly selected from the candidate list at the same time point t_k. The second design change consists of the interchange of two randomly selected design points $(\mathbf{x}_i, t_k)$ and $(\mathbf{x}_j, t_l)$ with $\mathbf{x}_i \neq \mathbf{x}_j$. The last design change concerns the deletion of a randomly chosen design point $\mathbf{x}_i$ at time point t_k and the addition of a design point $\mathbf{x}_j$ randomly selected from the candidate set at a not yet occupied and randomly chosen time point t_l. The design change leading to the largest positive effect on the optimality criterion (28) or (32) will be executed. This iterative improvement continues as long as a design change increases the criterion value. In order to avoid being stuck at a local optimum, the probability of finding the global optimum can be increased by repeating the search several times from different starting run orders or 'tries'. Additional avoidance of ending up in a local optimum can be obtained by applying simulated annealing techniques. Simulated annealing offers the ability to migrate through a sequence of local extrema in search of the global solution. Simulated annealing is extensively described in the next section and for a detailed outline

of the implemented design algorithm, we refer the interested reader to the appendix 3.

Finally, note that unlike conventional exchange algorithms, we can no longer make use of simple update formulas for the optimality criteria (28) and (32). This is due to the fact that these optimality criteria cannot be written as sums of outer products.

6.2 Simulated annealing

An excellent review of simulated annealing can be found in Rayward-Smith *et al.* (1996). The ideas that form the basis of simulated annealing date back to Metropolis *et al.* (1953). They developed an algorithm to simulate the change in energy of material in a heat batch when subjected to a cooling process, until it converges to a frozen state. For the behaviour of mechanical systems such as the one described by Metropolis *et al.* (1953), it is impossible to perform an exhaustive analysis of the possible energy states of the system. Consequently, one must have recourse to a statistical analysis. For instance, when atoms of a molten metal are cooled to a freezing temperature, they will tend to take relative positions in a lattice in such a way as to minimize the potential energy of their mutual forces. Because of the very large number of atoms and possible arrangements, the final state will most likely correspond to only a local energy minimum and not a global one. The solidified metal may be reheated and cooled slowly with the hope that it will then migrate to a lower energy state. In metallurgy this process is called annealing. In statistical mechanics, the probability that a system will transit from the state with energy e_1 to the state with higher energy e_2 is $\exp((e_1 - e_2)/kT)$ with k the Boltzmann constant and T the absolute temperature. The lower the temperature, the smaller the probability of transition to a higher energy state.

Kirkpatrick *et al.* (1983) and Cerny (1985) independently showed that the Metropolis algorithm could be applied to optimization problems by mapping the elements of the physical cooling process onto the elements of a combinatorial optimization problem. By this, simulated annealing could be used to find a global extremum of an objective function that has many local extrema. At the same time, it should be emphasized that the algorithm is a heuristic one and that the final optimum is not necessarily global. Firstly, an initial solution is randomly chosen or specified depending on available information. Suppose that the corresponding value of the objective function equals Q_1. Next, a neighbouring solution is randomly chosen. Suppose this neighbouring solution has function value Q_2. The resulting change in the value of the objective function then equals $Q_2 - Q_1$. In standard simulated annealing and for a function to be maximized, the move to the neighbouring solution is accepted with probability $\exp((Q_2 - Q_1)/T)$ if $Q_2 < Q_1$ and with probability 1 otherwise. The control parameter T acts as the temperature of the annealing process. This means that beneficial steps are accepted unconditionally but the detrimental steps are accepted according to an auxiliary experiment. A random number is generated from the uniform distribution on (0,1) and is compared to the value $\exp((Q_2 - Q_1)/T)$. If it is lower than $\exp((Q_2 - Q_1)/T)$, then the neighbouring solution is accepted during the next iteration, otherwise the neighbouring solution is rejected. By allowing moves to

inferior solutions, the chance of getting stuck in a poor local optimum is reduced. The temperature T is initially high, allowing many inferior moves to be accepted. In practice, one may choose the initial temperature so that after running the algorithm for a short time, an acceptance rate between 40% and 60% is observed. When an equilibrium is reached after a user-specified number of iterations at temperature T_i, the temperature is reduced to T_{i+1} according to a specified cooling schedule and the process is repeated. The iterations at temperature T_i are commonly referred to as plateau i and the number of iterations at plateau i is called the length of that plateau. As a rule of thumb, this length is chosen as a multiple of the neighbourhood size. Typically, the cooling schedule consists of a sequence of monotonically decreasing temperatures in order to reduce the probability of moving to a worse solution. A large number of variations in the choice of the cooling rate exists. The simplest one is geometric cooling where the temperature from plateau i to $i+1$ is reduced according to $T_{i+1} = \kappa T_i$, with the user-specified cooling rate κ typically in the range 0.85 to 0.95. The process of sequentially reducing the temperature continues until the final temperature is zero or until inferior moves are nearly always rejected.

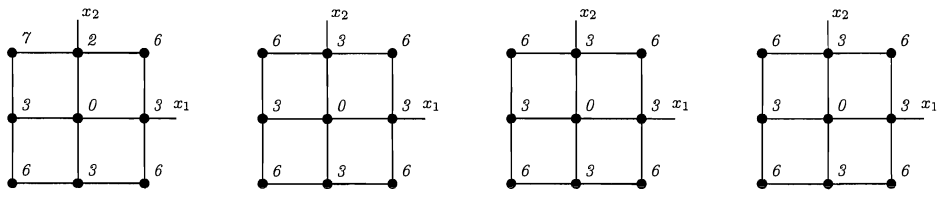
7 Parametric and semiparametric optimal designs

This section explicates the benefits of the application of semiparametric regression techniques in optimum design. Consider as an example the design of an experiment with 36 observations, the candidate set of design points taken from the 3^2 factorial and 36 equally spaced time points. The response model is represented by the polynomial expansion $\mathbf{f}'(\mathbf{x}) = [x_1, x_2, x_1x_2, x_1^2, x_2^2]$. Firstly, parametric $\mathcal{D}_t$-optimal run orders are computed for the following postulated time trends:

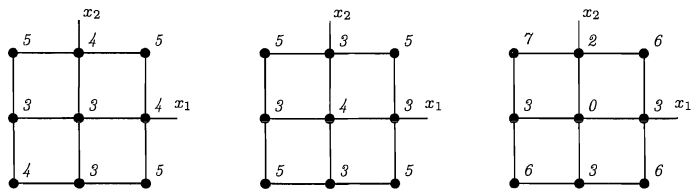
$$\begin{aligned} \mathbf{g}_1'(t) &= [t], \\ \mathbf{g}_2'(t) &= [t, t^2], \\ \mathbf{g}_3'(t) &= [t, t^2, t^3], \\ \mathbf{g}_4'(t) &= [t, t^2, t^3, t^4]. \end{aligned}$$

On the other hand, semiparametric $\mathcal{D}_t$-optimal run orders are computed for kernel function $K_1 = 1/2$ and for varying bandwidths between 0.1 and 1.0 and step size 0.1. The optimality criterion used to compute the run orders is based on partial residuals and is given in (32). The computed parametric and semiparametric $\mathcal{D}_t$-optimal designs mainly differ from each other in the number of replicates at the different design points. These numbers are shown in figure 1. The parametric design problems are referred to as PD, whereas the semiparametric design problems are indicated with the label SPD. The $\mathcal{D}$-optimal design is also depicted. For the parametric design problems with time trend $\mathbf{g}_2$, $\mathbf{g}_3$ or $\mathbf{g}_4$, the assignment of the observations to the different design points is the same. Two distinct assignments are obtained for the semiparametric design problems: one for the bandwidth equal to 0.1 and one for bandwidths larger than or equal to 0.2. From figure 1 one also observes that unlike the $\mathcal{D}$-optimal design and the parametric $\mathcal{D}_t$-optimal run

orders—for which all observations lie on the boundary of the design region—the semiparametric $\mathcal{D}_t$-optimal run orders have a number of observations at the center of the design region. This means that in case of semiparametric regression, the observations are spread more uniformly across the entire experimental region.



(a) PD, $\mathbf{g}_1$ (b) PD, $\mathbf{g}_2$ (c) PD, $\mathbf{g}_3$ (d) PD, $\mathbf{g}_4$



(e) SPD, $\lambda = 0.1$ (f) SPD, $\lambda \geq 0.2$ (g) $\mathcal{D}$-optimality

Figure 1: Number of replicates at the different design points of the parametric and semiparametric $\mathcal{D}_t$-optimal run orders

In order to compare both the bias and the precision of the parameter estimates $\hat{\boldsymbol{\alpha}}$, the parametric and the semiparametric $\mathcal{D}_t$-optimal run orders are compared to each other for the following five true time trends:

$$\begin{aligned}
\tilde{g}_1(t) &= 30t, \\
\tilde{g}_2(t) &= 30t^2, \\
\tilde{g}_3(t) &= -30t + 60t^3, \\
\tilde{g}_4(t) &= 150t^2 - 180t^4, \\
\tilde{g}_5(t) &= 30\sin(2\pi t).
\end{aligned}$$

A simulation study is performed with error terms drawn from the $N(0,1)$ distribution and the parameter vector $\boldsymbol{\alpha}$ equal to $[50, -30, 30, -15, 15]'$. For each $\mathcal{D}_t$-optimal run order and for each true time trend $\tilde{g}_i(t)$ with $i \in \{1, \ldots, 5\}$, we computed the euclidean distance between the true parameter vector $\boldsymbol{\alpha}$ and the average of the estimated parameters $\hat{\boldsymbol{\alpha}}$ over all 100,000 simulations. The bias results are shown in table 5. One easily sees that if the order of the true time trend $\tilde{g}(t)$ is larger than the order of the postulated time trend $\mathbf{g}(t)$, the parametric optimal run orders involve much more biased parameter estimates as compared to modeling the time trend nonparametrically. By this, modeling the time

dependence nonparametrically clearly outperforms a parametric fit when the order of the time trend is underspecified.

Table 5: Euclidean distances between the true parameter vector $\boldsymbol{\alpha}$ and the average of the estimated parameters $\hat{\boldsymbol{\alpha}}$ over all 100,000 simulations

	$\tilde{g}_1$	$\tilde{g}_2$	$\tilde{g}_3$	$\tilde{g}_4$	$\tilde{g}_5$
PD, $\mathbf{g}_1$	0.0021	9.4648	10.2994	11.2533	14.5530
PD, $\mathbf{g}_2$	0.0009	0.0009	8.2797	12.0224	8.7602
PD, $\mathbf{g}_3$	0.0021	0.0021	0.0021	12.6183	8.3857
PD, $\mathbf{g}_4$	0.0013	0.0013	0.0013	0.0013	16.8298
SPD, $\lambda = 0.1$	0.0624	0.5037	0.3076	3.4603	0.3955
SPD, $\lambda = 0.2$	0.2360	1.0118	1.1400	5.9583	1.5212
SPD, $\lambda = 0.3$	0.2045	0.7578	1.0983	5.0607	1.8243
SPD, $\lambda = 0.4$	0.5125	0.7284	2.3089	5.0045	6.0847
SPD, $\lambda = 0.5$	0.7115	1.2072	3.6678	7.0479	2.3271
SPD, $\lambda = 0.6$	0.8335	1.3651	2.2101	9.7911	8.0871
SPD, $\lambda = 0.7$	1.2761	0.7371	5.5284	3.9205	3.7472
SPD, $\lambda = 0.8$	0.9423	1.1078	5.9178	3.5715	4.5849
SPD, $\lambda = 0.9$	0.3956	2.1261	2.5921	1.7797	5.4625
SPD, $\lambda = 1.0$	0.2534	5.9866	1.3790	4.2145	4.6569

On the other hand, in terms of the precision of the important parameter estimates $\hat{\boldsymbol{\alpha}}$, the results are somewhat different. Table 6 displays the trend factors of the computed parametric and semiparametric run orders. The table reveals that the trend factors of the parametric run orders are considerably larger than those of the semiparametric run orders. This means that the outperformance of the semiparametric run orders in terms of the bias of the parameter estimates goes at the expense of the precision of the estimates. To combine both bias and precision, the parametric and the semiparametric run orders will be compared to each other with regard to the determinant of the mean squared error matrix

$$\mathbf{M} = \mathrm{E}\{(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})'\}. \tag{33}$$

For instance, element (i, j) of matrix $\mathbf{M}$ is equal to $\mathrm{E}\{(\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_j - \alpha_j)\} = \mathrm{covar}(\hat{\alpha}_i, \hat{\alpha}_j) + \mathrm{bias}(\hat{\alpha}_i)\mathrm{bias}(\hat{\alpha}_j)$. The computation of the matrix $\mathbf{M}$ is based on simulations in which the expectation is replaced by the average over all 100,000 simulation runs.

The results are displayed in table 7. It follows that in almost every case where the order of the time trend is underspecified, the determinant of $\mathbf{M}$ for the semiparametric run orders is smaller than that for the parametric run orders. Consequently, if the order of the postulated time trend is underspecified, one benefits from modeling the time dependence

Table 6: Trend factors

	TF
PD, $\mathbf{g}_1$	0.9999
PD, $\mathbf{g}_2$	0.9274
PD, $\mathbf{g}_3$	0.9273
PD, $\mathbf{g}_4$	0.8711
SPD, $\lambda = 0.1$	0.6365
SPD, $\lambda = 0.2$	0.7103
SPD, $\lambda = 0.3$	0.7207
SPD, $\lambda = 0.4$	0.7221
SPD, $\lambda = 0.5$	0.7233
SPD, $\lambda = 0.6$	0.7248
SPD, $\lambda = 0.7$	0.7257
SPD, $\lambda = 0.8$	0.7260
SPD, $\lambda = 0.9$	0.7262
SPD, $\lambda = 1.0$	0.7266

nonparametrically. It is very important to point out that this result is independent of the bandwidth chosen.

Summing up, if one has no certainty or knowledge about the form of the time trend or if the behaviour of the time trend is very hard to model parametrically, one benefits from modeling the time dependence nonparametrically. Table 7 has shown that in cases where the order of the parametric time trend is underspecified, modeling the time dependence nonparametrically clearly outperforms a purely parametric approach. This outperformance can be explained from the fact that in such cases the parametric run orders suffer more from the bias problem than the semiparametric run orders. An alternative way to guard against misspecification of the time trend may be to model a parametric time trend of high order. However, this approach is limited by an important computational consideration. In order to avoid singular design matrices, the number of observations and the number of distinct design points both have to be larger than the sum of the number of parameters in the response model and the trend model, i.e. $n \geq p + q$. For example, this means that for the full second-order factorial in two variables ($p = 6$) and nine observations ($n = 9$), the order q of the postulated time trend cannot be larger than three. Consequently, for true time trends of fourth order or more such as sines of high frequency for instance, the parametric $\mathcal{D}_t$-optimal run orders would still suffer from the bias problem. The next section illustrates the use of the simulated annealing algorithm for a real-life example.

Table 7: $10^7|\mathbf{M}|$ for the $\mathcal{D}_t$-optimal run orders

	$\tilde{g}_1$	$\tilde{g}_2$	$\tilde{g}_3$	$\tilde{g}_4$	$\tilde{g}_5$
PD, $\mathbf{g}_1$	4.36	4655	660	5634	11676
PD, $\mathbf{g}_2$	4.30	4.30	1350	7894	8121
PD, $\mathbf{g}_3$	4.29	4.29	4.29	9764	2304
PD, $\mathbf{g}_4$	4.77	4.77	4.77	4.77	40385
SPD, $\lambda = 0.1$	35	84	100	2467	143
SPD, $\lambda = 0.2$	39	177	544	5402	937
SPD, $\lambda = 0.3$	32	177	497	4806	1082
SPD, $\lambda = 0.4$	97	134	482	5494	1786
SPD, $\lambda = 0.5$	136	415	631	5612	1668
SPD, $\lambda = 0.6$	187	694	623	5712	2369
SPD, $\lambda = 0.7$	563	276	652	3651	2009
SPD, $\lambda = 0.8$	217	353	671	3567	2045
SPD, $\lambda = 0.9$	76	515	507	2948	2289
SPD, $\lambda = 1.0$	26	872	344	5114	1964

8 Reactor example

An industrial example is outlined to demonstrate the wide range of practical design problems that can be tackled with the presented design algorithm. An experiment is set up to evaluate the rate of conversion of a chemical reactor. After a screening experiment the experimenters found that the following variables are important to be investigated: feed rate x_1, temperature x_2, concentration x_3 and the amount of catalyst added x_4. The factors and the corresponding factor levels are shown in table 8. The assumed response model is represented by

$$\mathbf{f}'(\mathbf{x}) = [x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4, x_1^2, x_2^2].$$

Based on table 8, running the full factorial design would involve $3 \times 3 \times 2 \times 2 = 36$ factor level combinations but due to a restriction in the amount of raw material available, the number of observations is limited to 18. The design problem comes to efficiently selecting 18 out of the total number of 36 factor level combinations without replacement. The experimenters know that the reactor data may be influenced by an unknown time trend due to build-up of deposits on the reactor wall. Previous experience showed that the influence of deposits on the measurements is very complicated and extremely difficult to quantify. Therefore, the experimenters decide to guard against misspecification by modeling the time dependence nonparametrically.

The aim of this example is to demonstrate the influence of the estimation method, the kernel function and the bandwidth on the properties of the semiparametric $\mathcal{D}_t$-optimal run orders. For varying bandwidths between 0.2 and 1.0 with step size 0.2 and the kernel

Table 8: Factors and factor levels in the reactor experiment

factor	levels	coded levels
feed rate x_1 (liters/min)	50, 70, 90	-1, 0, 1
temperature x_2 (degrees Celsius)	200, 230, 260	-1, 0, 1
concentration x_3 (%)	5, 8	-1, 1
catalyst x_4 (%)	0.5, 2	-1, 1

functions of table 4, the design algorithm is used to compute run orders that maximize optimality criterion (28) or (32). The computed semiparametric $\mathcal{D}_t$-optimal run orders contain observations at almost every corner point of the form $(\pm 1, \pm 1, \pm 1, \pm 1)$. This means that design points on the corners of the experimental region have the highest chance of being selected by the semiparametric $\mathcal{D}_t$-optimality criterion.

To compare the semiparametric run orders to a purely parametric fitting, a simulation study is conducted with $\boldsymbol{\alpha}$ equal to $[10, -10, 10, 20, 30, -20, 20, -10, -10, -40, -30, 50]'$, time trend $g(t) = 50\sin(\pi t)$ and error terms drawn from the $N(0, 1)$ distribution. For the parametric run orders computed with a first-order and a second-order polynomial time trend, $10^{10}|\mathbf{M}|$ is equal to 2.95 and 8.70 respectively. The results for the semiparametric run orders are shown in table 9.

Table 9: $10^{10}|\mathbf{M}|$ for the semiparametric $\mathcal{D}_t$-optimal run orders of the reactor experiment

	Green-Jennison-Seheult					
λ	K_1 $(m=2)$	K_2 $(m=4)$	K_3 $(m=6)$	K_4 $(m=2)$	K_5 $(m=4)$	K_6 $(m=6)$
0.2	1.51	1.71	2.35	1.95	1.98	3.11
0.4	2.62	1.34	1.85	2.84	1.63	1.88
0.6	1.55	1.33	2.27	2.01	1.94	1.78
0.8	1.88	1.26	3.40	1.65	2.14	2.96
1.0	1.66	2.08	1.08	1.32	1.95	2.10
	partial residuals					
λ	K_1 $(m=2)$	K_2 $(m=4)$	K_3 $(m=6)$	K_4 $(m=2)$	K_5 $(m=4)$	K_6 $(m=6)$
0.2	1.61	1.42	1.45	2.00	1.42	2.24
0.4	2.12	1.94	1.14	2.36	2.25	1.22
0.6	1.41	1.63	2.38	1.97	1.42	1.85
0.8	1.68	2.33	3.14	1.09	2.22	1.25
1.0	3.55	2.11	1.55	1.99	1.25	2.58

The simulation study reveals that the computed semiparametric run orders outperform the parametric run orders for almost every combination of bandwidth, kernel function

and estimation method. By this, the fact that the semiparametric run orders outperform the parametric ones is independent of the estimation method, the kernel function and the bandwidth chosen. The simulation study does not show an obvious pattern in the influence of this choice on the determinant of the mean squared error matrix $\mathbf{M}$. However, studying the bias properties and the precision of the parameter estimates separately, truly leads to some useful insights.

Table 10 shows the trend factor for each semiparametric $\mathcal{D}_t$-optimal run order. Some important conclusions can be drawn. Firstly, the table reveals that for a given bandwidth and kernel function, in almost every case the $\mathcal{D}_t$-optimal run order for the Green-Jennison-Seheult criterion (28) has a larger trend factor than the corresponding run order computed with the criterion based on partial residuals. Besides, it holds in general that the larger the bandwidth, the higher the trend factor of the optimal run order. Alternatively stated, the generalized variance of the important parameter estimates decreases as the bandwidth grows larger. This observation is in accordance with the results of table 6 in section 7 and a similar conclusion was obtained by Speckman (1988) in a regression context instead of a design context. Thirdly, for a given bandwidth, a minimum variance kernel nearly always involves a larger trend factor than when the optimal kernel of the same order is used. Finally, the low order kernels K_1 and K_4 lead in almost every case to the largest values of the trend factor. We conclude that in terms of the $\mathcal{D}_t$-optimality criterion, the Green-Jennison-Seheult criterion (28), high bandwidths and minimum variance kernels of low order have to be preferred. For instance, the $\mathcal{D}_t$-optimal run order for the Green-Jennison-Seheult criterion, $\lambda = 1.0$ and kernel function K_1 has trend factor 0.771, whereas the trend factor of the $\mathcal{D}_t$-optimal run order for criterion (32), $\lambda = 0.2$ and kernel function K_6 is 0.663.

However, one has to be careful with following the directives made in terms of the generalized variance. It is also important to investigate the bias properties of the semiparametric run orders. For each $\mathcal{D}_t$-optimal run order, the Euclidean distance between the parameter vector $\boldsymbol{\alpha}$ and the average parameter estimates $\hat{\boldsymbol{\alpha}}$ over all 100,000 simulations is calculated. The results are shown in table 11. Whereas in terms of $\mathcal{D}_t$-optimality, the Green-Jennison-Seheult criterion and large bandwidths are recommended, it can clearly be seen that, in general, for a given kernel function the optimal run order for the Green-Jennison-Seheult criterion and $\lambda = 1.0$ suffers more from the bias problem than the optimal run orders for the criterion based on partial residuals and a small bandwidth. Besides, for a given estimation method and a particular bandwidth, the bias of the optimal run order computed with the minimum variance kernel of low order, namely K_1, is in general larger than the bias of the optimal run order computed with the optimal kernel of the highest order, namely K_6. It follows from the bias properties that the preference for the Green-Jennison-Seheult criterion with a large bandwidth and a low order minimum variance kernel has to be weakened in favour of the estimation method based on partial residuals and small bandwidths.

Summing up, the outperformance of the semiparametric run orders is independent of the choice of the estimation method, the bandwidth and the kernel function. The aim of

Table 10: Trend factor for the semiparametric $\mathcal{D}_t$-optimal run orders of the reactor experiment

	Green-Jennison-Seheult					
λ	K_1 $(m=2)$	K_2 $(m=4)$	K_3 $(m=6)$	K_4 $(m=2)$	K_5 $(m=4)$	K_6 $(m=6)$
0.2	0.718	0.722	0.720	0.718	0.717	0.714
0.4	0.764	0.710	0.657	0.754	0.670	0.710
0.6	0.770	0.729	0.740	0.761	0.734	0.678
0.8	0.771	0.766	0.725	0.771	0.754	0.722
1.0	0.771	0.769	0.749	0.771	0.761	0.739
	partial residuals					
λ	K_1 $(m=2)$	K_2 $(m=4)$	K_3 $(m=6)$	K_4 $(m=2)$	K_5 $(m=4)$	K_6 $(m=6)$
0.2	0.675	0.665	0.670	0.666	0.668	0.663
0.4	0.748	0.661	0.661	0.737	0.648	0.667
0.6	0.767	0.704	0.684	0.758	0.696	0.644
0.8	0.769	0.755	0.695	0.768	0.730	0.681
1.0	0.770	0.760	0.732	0.771	0.757	0.728

this section was to investigate how this choice influences the balance between bias and precision of the outperforming semiparametric run orders. Using the Green-Jennison-Seheult estimation method with a large bandwidth and a minimum variance kernel of low order involves an increased precision but goes at the expense of the bias. On the other hand, applying the estimation method based on partial residuals, using a small bandwidth and an optimal kernel leads to reduced bias and goes at the cost of the precision of the important parameter estimates. The balance desired depends on the design problem at hand and must be chosen by the experimenter himself.

9 Conclusion

This paper has presented an extension of the $\mathcal{D}_t$-optimality criterion to design problems in which the response is modeled semiparametrically. A design algorithm is provided for the construction of run orders that are optimally balanced for nonparametric time trends. The advantage of modeling the time dependence nonparametrically lies in the fact that it outperforms a full parametric approach in terms of the combined effect of bias and precision of the important parameter estimates when the parametric time trend is underspecified. The outperformance of the semiparametric run orders is independent of the estimation method, the bandwidth and the kernel function chosen. This choice only influences the balance between bias and precision of the outperforming semiparametric run orders. For instance, preferring the Green-Jennison-Seheult estimators with a high bandwidth value and a minimum variance kernel of low order benefits the precision of

Table 11: Euclidean distance for the semiparametric $\mathcal{D}_t$-optimal run orders of the reactor experiment

	Green-Jennison-Seheult					
λ	K_1 $(m=2)$	K_2 $(m=4)$	K_3 $(m=6)$	K_4 $(m=2)$	K_5 $(m=4)$	K_6 $(m=6)$
0.2	19.9	22.7	25.4	8.2	7.8	25.1
0.4	25.9	12.6	17.9	20.0	10.2	23.3
0.6	14.2	11.4	25.4	27.5	20.1	9.7
0.8	22.5	27.9	7.7	23.0	24.8	14.1
1.0	28.0	18.3	24.2	28.1	22.9	23.9
	partial residuals					
λ	K_1 $(m=2)$	K_2 $(m=4)$	K_3 $(m=6)$	K_4 $(m=2)$	K_5 $(m=4)$	K_6 $(m=6)$
0.2	12.9	5.6	10.0	11.4	9.0	6.6
0.4	26.1	12.3	22.0	6.4	10.4	4.7
0.6	15.6	11.2	16.2	23.3	9.9	5.2
0.8	18.4	19.2	5.2	23.5	11.3	9.6
1.0	26.3	19.4	7.6	11.3	23.4	11.8

the parameter estimates and goes at the expense of the parameter bias. On the contrary, reduced biased is obtained by using the estimation method based on partial residuals, small bandwidth values and optimal kernels. The choice is up to the experimenter and depends on his or her preference.

Appendix 1: Parameter bias for $q = q_1$

For the statistical model

$$\mathbf{y} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{G}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}, \tag{1.1}$$

the least-squares estimators $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}_1$ are found by minimizing the quantity

$$(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\alpha}} - \mathbf{G}\hat{\boldsymbol{\beta}}_1)'(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\alpha}} - \mathbf{G}\hat{\boldsymbol{\beta}}_1). \tag{1.2}$$

Setting the derivative of (1.2) with respect to $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}_1$ equal to zero, gives the normal equations

$$\begin{aligned} \mathbf{F}'\mathbf{F}\hat{\boldsymbol{\alpha}} &= \mathbf{F}'(\mathbf{y} - \mathbf{G}_1\hat{\boldsymbol{\beta}}), \\ \mathbf{G}_1'\mathbf{G}_1\hat{\boldsymbol{\beta}}_1 &= \mathbf{G}_1'(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\alpha}}). \end{aligned}$$

It is handy to rewrite the normal equations for $\hat{\boldsymbol{\beta}}_1$ as

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\alpha}}). \tag{1.3}$$

Substituting (1.3) in the normal equations for $\hat{\boldsymbol{\alpha}}$ gives

$$\hat{\boldsymbol{\alpha}} = [\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{F}]^{-1}\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{y}. \tag{1.4}$$

If the true regression model is given by

$$\mathbf{y} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{G}_1\boldsymbol{\beta}_1 + \mathbf{G}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \tag{1.5}$$

then the expected value of the parameter estimates (1.4) is equal to

$$\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{\alpha}}) &= [\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{F}]^{-1}\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathrm{E}(\mathbf{y}),\\
&= [\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{F}]^{-1}\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathrm{E}(\mathbf{F}\boldsymbol{\alpha} + \mathbf{G}_1\boldsymbol{\beta}_1 + \mathbf{G}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}),\\
&= [\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{F}]^{-1}\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}(\mathbf{F}\boldsymbol{\alpha} + \mathbf{G}_1\boldsymbol{\beta}_1 + \mathbf{G}_2\boldsymbol{\beta}_2 + \mathrm{E}(\boldsymbol{\varepsilon})),\\
&= [\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{F}]^{-1}\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}(\mathbf{F}\boldsymbol{\alpha} + \mathbf{G}_1\boldsymbol{\beta}_1 + \mathbf{G}_2\boldsymbol{\beta}_2),\\
&= \boldsymbol{\alpha} + [\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{F}]^{-1}\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{G}_1\boldsymbol{\beta}_1\\
&\quad + [\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{F}]^{-1}\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{G}_2\boldsymbol{\beta}_2,\\
&= \boldsymbol{\alpha} + [\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{F}]^{-1}\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{G}_2\boldsymbol{\beta}_2.
\end{aligned}$$

By this, the parameter bias equals

$$\mathrm{E}(\hat{\boldsymbol{\alpha}}) - \boldsymbol{\alpha} = [\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{F}]^{-1}\mathbf{F}'\{\mathbf{I} - \mathbf{G}_1(\mathbf{G}_1'\mathbf{G}_1)^{-1}\mathbf{G}_1'\}\mathbf{G}_2\boldsymbol{\beta}_2. \tag{1.6}$$

Appendix 2: Parameter bias for $q = q_1 + q_2$

For the statistical model

$$\mathbf{y} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{G}_1\boldsymbol{\beta}_1 + \mathbf{G}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{G}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \tag{2.1}$$

with $\mathbf{Z} = [\mathbf{F}\ \mathbf{G}_1]$ and $\boldsymbol{\gamma} = [\boldsymbol{\alpha}'\ \boldsymbol{\beta}_1']'$, the least-squares estimators are given by

$$\begin{aligned}
\begin{bmatrix} \hat{\boldsymbol{\gamma}} \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} &= \left\{ \begin{bmatrix} \mathbf{Z}' \\ \mathbf{G}_2' \end{bmatrix} [\mathbf{Z}\ \mathbf{G}_2] \right\}^{-1} \begin{bmatrix} \mathbf{Z}' \\ \mathbf{G}_2' \end{bmatrix} \mathbf{y},\\
&= \begin{bmatrix} \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{G}_2 \\ \mathbf{G}_2'\mathbf{Z} & \mathbf{G}_2'\mathbf{G}_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Z}' \\ \mathbf{G}_2' \end{bmatrix} \mathbf{y}.
\end{aligned} \tag{2.2}$$

If the true regression model is

$$\mathbf{y} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{G}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{2.3}$$

then the expected value of the parameter estimates (2.2) is equal to

$$\begin{aligned}
\mathrm{E}\left(\begin{bmatrix} \hat{\boldsymbol{\gamma}} \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} \right) &= \begin{bmatrix} \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{G}_2 \\ \mathbf{G}_2'\mathbf{Z} & \mathbf{G}_2'\mathbf{G}_2' \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Z}' \\ \mathbf{G}_2' \end{bmatrix} \mathrm{E}(\mathbf{y}),\\
&= \begin{bmatrix} \mathbf{Z}'\mathbf{Z} & \mathbf{Z}'\mathbf{G}_2 \\ \mathbf{G}_2'\mathbf{Z} & \mathbf{G}_2'\mathbf{G}_2' \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Z}'\mathbf{Z}\boldsymbol{\gamma} \\ \mathbf{G}_2'\mathbf{Z}\boldsymbol{\gamma} \end{bmatrix}.
\end{aligned} \tag{2.4}$$

Based on theorem 8.5.11 of Harville (1997), it follows that

$$\mathrm{E}\left(\begin{bmatrix} \hat{\boldsymbol{\gamma}} \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix}\right) = \begin{bmatrix} (\mathbf{Z}'\mathbf{Z})^{-1} + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{G}_2\mathbf{Q}^{-1}\mathbf{G}_2'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} & -(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{G}_2\mathbf{Q}^{-1} \\ -\mathbf{Q}^{-1}\mathbf{G}_2'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} & \mathbf{Q}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{Z}'\mathbf{Z}\boldsymbol{\gamma} \\ \mathbf{G}_2'\mathbf{Z}\boldsymbol{\gamma} \end{bmatrix}, \tag{2.5}$$

with $\mathbf{Q} = \mathbf{G}_2'\mathbf{G}_2 - \mathbf{G}_2'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{G}_2$. It can now easily be verified that

$$\mathrm{E}(\hat{\boldsymbol{\gamma}}) = \boldsymbol{\gamma} \tag{2.6}$$

and

$$\mathrm{E}(\hat{\boldsymbol{\beta}}_2) = \mathbf{0}. \tag{2.7}$$

Expression (2.6) also implies that

$$\mathrm{E}(\hat{\boldsymbol{\alpha}}) - \boldsymbol{\alpha} = \mathbf{0}. \tag{2.8}$$

Consequently, the parameter estimates $\hat{\boldsymbol{\alpha}}$ are unbiased when the order of the time trend is overspecified.

Appendix 3: The design algorithm

In the outline of the algorithm, the list of d candidate design points is written as $D = \{\mathbf{x}_1, \ldots, \mathbf{x}_d\}$ and the list of h available time points is given as $H = \{t_1, \ldots, t_h\}$. A run order is described by a series $R = \{(\mathbf{x}_i, t_j)\}$ and the corresponding criterion value (28) or (32) is denoted as Q. For instance, the initial run order is denoted as R_0 with criterion value Q_0. γ denotes the number of the plateau, μ the length of each plateau and ω the user-specified number of plateaus. This number ω has to be chosen such that the probability of accepting a detrimental move becomes almost negligible for the final plateau. R_γ denotes the run order at the end of plateau γ and the corresponding criterion value equals Q_γ. Besides, the number of initial run orders or tries equals ν and υ is a user-specified number determining the stopping criterion after the final plateau ω has been reached. The stopping criterion is satisfied if no further significant improvement of the run order is obtained.

The input to the algorithm consists of the number of observations n, the number of factors, the polynomial expansion $\mathbf{f}(\mathbf{x})$ of the response model, the list of candidate points D with potential restrictions on the number of replicates per design point, the list of time points H, the estimation method, the kernel function K, the bandwidth λ, the number of starting designs ν, the stopping criterion υ, the initial temperature, the number of plateaus ω, the length of each plateau μ and the cooling rate κ. After reading the input, the standard simulated annealing algorithm with geometric cooling proceeds as follows:

1. Set $Q_{\text{opt}} = 0$ and $R_{\text{opt}} = \{\}$.

2. Repeat ν times:

   (a) Set $Q = 0$ and $R = \{\}$.

   (b) Repeat n times:

      i. Randomly choose $\mathbf{x}_i \in D$.

      ii. Randomly choose $t_k \in H$.

      iii. Set $R = R \cup \{(\mathbf{x}_i, t_k)\}$.

      iv. Set $H = H \setminus \{t_k\}$.

   (c) Update Q.

   (d) Set $Q_0 = Q$ and $R_0 = R$.

   (e) Set $\gamma = 1$.

   (f) Repeat μ times:

      i. Select a random design change (1):
      Randomly choose $(\mathbf{x}_i^{(1)}, t_k^{(1)}) \in R$ and $\mathbf{x}_j^{(1)} \in D$ with $\mathbf{x}_i^{(1)} \neq \mathbf{x}_j^{(1)}$:
      compute the relative effect $\triangle^{(1)}$ on Q of deleting $(\mathbf{x}_i^{(1)}, t_k^{(1)})$ and adding $(\mathbf{x}_j^{(1)}, t_k^{(1)})$.

      ii. Select a random design change (2):
      Randomly choose $(\mathbf{x}_i^{(2)}, t_k^{(2)})$ and $(\mathbf{x}_j^{(2)}, t_l^{(2)}) \in R$ with $\mathbf{x}_i^{(2)} \neq \mathbf{x}_j^{(2)}$:
      compute the relative effect $\triangle^{(2)}$ on Q of interchanging $(\mathbf{x}_i^{(2)}, t_k^{(2)})$ and $(\mathbf{x}_j^{(2)}, t_l^{(2)})$.

      iii. Select a random design change (3):
      Randomly choose $(\mathbf{x}_i^{(3)}, t_k^{(3)}) \in R$, $\mathbf{x}_j^{(3)} \in D$ and $t_l^{(3)} \in H$:
      compute the relative effect $\triangle^{(3)}$ on Q of deleting $(\mathbf{x}_i^{(3)}, t_k^{(3)})$ and adding $(\mathbf{x}_j^{(3)}, t_l^{(3)})$.

      iv. Set $\triangle = \max\{\triangle^{(1)}, \triangle^{(2)}, \triangle^{(3)}\}$.

      v. If $(\triangle \leq 1)$ then generate a random number $\xi \in (0, 1)$.

      vi. If $((\triangle > 1)$ or $((\triangle \leq 1)$ and $(\xi < \exp((1 - \triangle^{-1})\tau^{-1}))))$ then

         A. $Q = \triangle Q$

         B. If $\triangle = \triangle^{(1)}$ then $R = R \setminus \{(\mathbf{x}_i^{(1)}, t_k^{(1)})\} \cup \{(\mathbf{x}_j^{(1)}, t_k^{(1)})\}$.

         C. If $\triangle = \triangle^{(2)}$ then $R = R \setminus \{(\mathbf{x}_i^{(2)}, t_k^{(2)}), (\mathbf{x}_j^{(2)}, t_l^{(2)})\} \cup \{(\mathbf{x}_i^{(2)}, t_l^{(2)}), (\mathbf{x}_j^{(2)}, t_k^{(2)})\}$.

         D. If $\triangle = \triangle^{(3)}$ then $R = R \setminus \{(\mathbf{x}_i^{(3)}, t_k^{(3)})\} \cup \{(\mathbf{x}_j^{(3)}, t_l^{(3)})\}$ and $H = H \setminus \{t_k^{(3)}\} \cup \{t_l^{(3)}\}$.

   (g) Set $Q_\gamma = Q$ and $R_\gamma = R$.

   (h) If $((\gamma < \omega)$ or $((\gamma \geq \omega)$ and $(Q_\gamma > (1 + \upsilon)Q_{\gamma-1})))$ then

      i. Set $\tau = \kappa\tau$.

      ii. Set $\gamma = \gamma + 1$.

      iii. go to step (g)

   (i) If $Q > Q_{\text{opt}}$ then

      i. Set $Q_{\text{opt}} = Q$.

      ii. Set $R_{\text{opt}} = R$.

3. Write Q_{opt} and R_{opt}.

The algorithm is implemented in Fortran and uses the random number generator 'rand' of the Netlib library of Bell Labs and the NAG Fortran Library routine 'f03aaf' to calculate the determinant of a matrix. The output of the algorithm consists of the semiparametric $\mathcal{D}_t$-optimal run order R_{opt} with corresponding criterion value Q_{opt}.

References

Atkinson, A. C. and Donev, A. N. (1996), "Experimental Designs Optimally Balanced for Trend," *Technometrics*, **38**, 333-341.

Box, G. E. P. and Behnken, D. W. (1960), "Some New Three Level Designs for the Study of Quantitative Variables," *Technometrics*, **2**, 455-475.

Box, G. E. P. and Wilson, K. B. (1951), "On the Experimental Attainment of Optimum Conditions," *Journal of the Royal Statistical Society, Series B*, **13**, 1-38.

Box, M. J. and Draper, N. R. (1971), "Factorial Designs, the $|\mathbf{X}'\mathbf{X}|$ Criterion and Some Related Matters," *Technometrics*, **13**, 731-742.

Bradley, R. A. and Yeh, C.-M. (1980), "Trend-Free Block Designs: Theory," *The Annals of Statistics*, **8**, 883-893.

Butler, N. A. (1989), "When Optimal Designs for Polynomial Regression are Optimal for Smoothing Spline Models," *Technical Report 98/3*, University of Reading.

Butler, N. A. (1990), "On Weighted Design Optimality Criteria for Polynomial and Smoothing Spline Response Surfaces," *Technical Report 99/2A*, University of Reading.

Cerny, V. (1985), "A Thermodynamical Approach to the Travelling Salesman Problem: An Efficient Simulation Algorithm," *Journal of Optimization Theory and Applications*, **45**, 41-52.

Clark, R. M. (1977), "Nonparametric Estimation of a Smooth Regression Function," *Journal of the Royal Statistical Society, Series B*, **39**, 107-113.

Craven, P. and Wahba, G. (1979), "Smoothing Noisy Data With Spline Functions," *Numerical Mathematics*, **31**, 377-403.

Epanechnikov, V. A. (1969), "Nonparametric Estimation of a Multivariate Probability Density," *Theory of Probability and Its Applications*, **14**, 153-158.

Faraway, J. (1990), "Sequential Design for the Nonparametric Regression of Curves and Surfaces," *Technical Report 177*, University of Michigan.

Faraway, J. and Rothman, E. (1989), "Sequential Design for Nonparametric Regression," *Technical Report 174*, University of Michigan.

Gasser, T., Müller, H. G. and Mammitzsch, V. (1985), "Kernels for Nonparametric Curve Estimation," *Journal of the Royal Statistical Society, Series B*, 47, 238-252.

Green, P., Jennison, C. and Seheult, A. (1985), "Analysis of Field Experiments by Least Squares Smoothing," *Journal of the Royal Statistical Society, Series B*, 47, 299-315.

Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge University Press, New York.

Harville, D. A. (1997), *Matrix Algebra From A Statistician's Perspective*, Springer-Verlag, New York.

Kirkpatrick, S., Gelatt, C. D., Jr., and Vechhi, M. P. (1983), "Optimization by Simulated Annealing," *Science*, **220**, 671-680.

Metropolis, N., Rosenbluth, A. W., Teller, A. H. and Teller, E. (1953), "Equation of State Calculation by Fast Computing Machines," *Journal of Chemical Physics*, **21**, 1087-1091.

Müller, H. (1984), "Optimal Designs for Nonparametric Regression," *Statistics and Probability Letters*, **2**, 285-290.

Rayward-Smith, V. J., Osman, I. H., Reeves, C. R. and Smith, G. D. (1996). *Modern Heuristic Search Methods.* John Wiley and Sons, Chichester.

Rice, J. (1984), "Boundary Modification for Kernel Regression," *Communications in Statistics*, **13**, 893-900.

Speckman, P. (1988), "Kernel Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society, Series B*, **50**, 413-446.

Tack, L. and Vandebroek, M. (2000a), "Trend-resistant and Cost-efficient Fixed or Random Block Designs," *Under revision.*

Tack, L. and Vandebroek, M. (2000b), "Trend-resistant Designs under Budget Constraints," *Submitted for publication.*

Tack, L. and Vandebroek, M. (2001), "$(\mathcal{D}_t, C)$-Optimal Run Orders," *To appear in Journal of Statistical Planning and Inference.*