第33卷第3期 上海师范大学学报(自然科学版) Vol. 33, No. 3
2004年9月 Journal of Shanghai Normal University (Natural Sciences) Sep. 2004

Robust designs for approximate regression models with correlated errors

ZHOU Xiao-dong, YUE Rong-xian

(Mathematics and Sciences College, Shanghai Normal University, Shanghai 200234, China)

Abstract: This paper studies the model - robust design problem by applying the reproducing kernel space approach. We assume that the model has an unknown bias or contamination from some class $\mathscr{H}$ with a probability measure P, and the correlated errors are considered. We develop a design criterion in terms of the average expected quadratic loss for generalized least squares estimation. Numerical results indicate the designs obtained through this criterion are more robust.

Key words: Robust design; reproducing kernel Hilbert space; correlated error

CLC number: O211.6 **Document code**: A **Article ID**: 1000-5137(2004)03-0016-07

1 Introduction

In this paper we study the design problem for the linear regression given by

$$y_i = \sum_{j=1}^{p} \theta_j g_j(\mathbf{x}_i) + \varepsilon_i, \ i = 1, \cdots, n \tag{1}$$

when there is a deviation (misspecification or bias) from the assumed model. Here the specified functions g_i are linearly independent, and $\mathbf{x}_i$ are n points drawn from a compact subset in the q - dimensional Euclidean space. Differing from the classical assumption, we suppose errors ε_i, with mean zero, are not independent, and the covariance matrix is $\sigma^2\Sigma$. We represent the true model by

$$y_i = f^T(\mathbf{x}_i)\theta + \tau h(\mathbf{x}_i) + \varepsilon_i, \ i = 1, \cdots, n. \tag{2}$$

where $f(\cdot) = (g_1(\cdot), \cdots, g_p(\cdot))^T$ and $\mathscr{H}$ is an unknown function from some class $\mathscr{H}$ which will be specified later. The parameter τ, which only depends on n, reflects the relationship of the bias $\mathscr{H}$ and the error ε, as taken by Wiens and Zhou (1999). Since the bias $\mathscr{H}$ is unknown and may vary freely in H, the designs must be chosen such that the fitted model provides an adequate approximation to a range of possible true models, i. e., is robust to the exact form of the true model in some sense.

The model - robust designs problem has been studied by many authors whose investigations differ in specifi-

Received date: 2004-04-10

Foundation item: This work is partially supported by the National Science Foundation of China grant 10271078, and the Special Funds for Major Specialties of the Shanghai Education Committee and Shanghai Normal University.

Biography: ZHOU Xiao-dong (1979-), male, postgraduate, Mathematics and Sciences College, Shanghai Normal University. YUE Rong-xian (1958-), male, professor, Mathematics and Sciences College, Shanghai Normal University.

cation of the class H, the design region, the regressor, the loss functions and the covariance matrix. Kiefer (1973) and some others restrict their attentions to finite dimensional $\mathscr{H}$ and the least square estimators or linear estimators. Pesotchinsky(1982), Li and Notz(1982) and Yue and Hickernell(1999) deal with infinite dimensional H. Under the assumption of homoscedasticity, some of them take $H = \{h: |h(\mathbf{x})| \leqslant \phi(\mathbf{x}), \mathbf{x} \in X\}$ with various assumptions being made about $\phi(\mathbf{x})$. The designs constructed appear to be quite sensitive to the assured form of ϕ, The others take $H = \{h: \int_X [h(\mathbf{x})]^2 d\mathbf{x} \leqslant \eta, \int_X g_i(\mathbf{x})h(\mathbf{x})d\mathbf{x} = 0, j = 1, \cdots, p\}$, and use the least squares estimators. Here η is assumed to be known, and the second condition ensures the identifiability of the θ_j. But this specification is criticized for only designs which are absolutely continuous on X have a finite loss. To avoid this limitation, Yue and Hickernell(1999) allow for $\mathscr{H}$ coming from a reproducing kernel Hilbert space admitting a reproducing kernel $K(\mathbf{x}, \mathbf{w})$ and an inner product $\langle \cdot, \cdot \rangle$.

The aforementioned models are all supposed to be with homoscedasticity, but this rarely happens in practice. In this paper, we consider the model with correlated errors and H, the bias class, is a probability space with measure P. We confine ourselves to using the generalized least squares and the average expected quadratic loss. The next section gives the formulation of the problem along with the required notation. In Section 3, we give several examples to illustrate the design problem and also investigate the effect caused by the correlation of the errors. In Section 4 a summary is given.

2 Formulation of the design problem

Assume that the true response is given by (2). Using matrix notation, we can write the model (2) as

$$Y = X\theta + \tau H + \varepsilon. \tag{3}$$

where

$$Y = (y_1, \cdots, y_n)^T, \theta = (\theta_1, \cdots, \theta_p)^T,$$
$$H = (h(\mathbf{x}_1), \cdots, h(\mathbf{x}_n))^T, \varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^T.$$

and X is an $n \times p$ matrix of full column rank whose ith row contains the elements of $f^T(\mathbf{x}_i)$ $(i = 1, \cdots, n)$. From the assumption on the random errors in (2) we have $E[\varepsilon] = 0$, $Cov[\varepsilon] = \sigma^2\Sigma$, where $\Sigma > 0$ is an $n \times n$ matrix.

We use the best linear unbiased estimator to estimate the parameters' vector θ when the contamination is not present. That is, $\hat{\theta} = CY$ where $C = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}$. The vector for predicted response, $\hat{y}(\mathbf{x})$, can be expressed by $\hat{y}(\mathbf{x}) = f^T(\mathbf{x})\hat{\theta}$. Let $\eta(\mathbf{x})$ be defined as corresponding true mean values at the point $\mathbf{x}$. When the contamination $h(\mathbf{x})$ is present, we have

$$E[\hat{y}(\mathbf{x})] = \eta(\mathbf{x}) + \tau f^T(\mathbf{x})CH - \tau h(\mathbf{x}),$$
$$Var[\hat{y}(\mathbf{x})] = \sigma^2 f^T(\mathbf{x})(X^T\Sigma^{-1}X)^{-1}f(\mathbf{x}).$$

We consider the expected quadratic loss over the region X when the design ξ_n and estimator $\hat{\theta}$ are used

$$R(\xi_n, h) = \int_X E(\hat{y}(\mathbf{x}) - \eta(\mathbf{x}))^2 d\mathbf{x}. \tag{4}$$

Introduce the matrices $M = X^T\Sigma^{-1}X$, $\Gamma = \int_X f(\mathbf{x})f^T(\mathbf{x})d\mathbf{x}$. Then (4) becomes

$$R(\xi_n, h) = \sigma^2 tr[M^{-1}\Gamma] + \tau^2\{tr[C^T\Gamma CHH^T] - 2\int_X H^TC^Tf(\mathbf{x})h(\mathbf{x})d\mathbf{x} + \int_X h^2(\mathbf{x})d\mathbf{x}\}. \tag{5}$$

Note that the bias $h(\mathbf{x})$ is unknown, and we must make an assumption before it can make sense to minimize (5). Let us suppose that the class $\mathscr{H}$ is a probability space with measure P. Equivalently, the bias $\mathscr{H}$ is regarded as a random function and P represents the probability that $\mathscr{H}$ presents. We additionally assume that

$$E[h(\mathbf{x})] = \int_{\mathscr{H}} h(\mathbf{x})\, dP(h) = 0,\ E[h(\mathbf{x})h(\mathbf{t})] = K(\mathbf{x}, \mathbf{t}),$$

where K is a specified covariance kernel. The assumption $E[h(\mathbf{x})] = 0$ reflects the notion that the model in (2) is correct on the average, but any particular realization may induce the bias. Our objective will be to find ξ_n to minimize the average expected quadratic loss

$$\bar{R}(\xi_n) = \int_{\mathscr{H}} R(\xi_n, h)\, dP(h). \tag{6}$$

Let $K(\mathbf{x})$ be the $n \times 1$ vector defined by $K(\mathbf{x}) = (K(\mathbf{x},\mathbf{x}_1), \cdots, K(\mathbf{x},\mathbf{x}_n))^T$, let Kbe the $n \times n$ matrix whose i th row is $K^T(\mathbf{x}_i)$,for $i = 1,\cdots,n$. Define the matrix L and the constant d by $L = \int_{\chi} f(\mathbf{x})K^T(\mathbf{x})\,d\mathbf{x}$ and $d = \int_{\chi} K(\mathbf{x},\mathbf{x})\,d\mathbf{x}$. Then $\bar{R}(\xi_n)$ in (6) can be expressed as

$$\bar{R}(\xi_n) = \sigma^2 tr[M^{-1}\Gamma] + \tau^2\{tr[C^T\Gamma CK] - 2tr[C^TL] + d\}. \tag{7}$$

Let $\lambda = \frac{\sigma^2}{\sigma^2 + \tau^2}$, which can be chosen by the experimenter to reflect his view of the relative importance of variance versus bias. Then (7) becomes

$$\begin{aligned}\bar{R}(\xi_n) &= (\sigma^2 + \tau^2)\{\lambda tr[M^{-1}\Gamma] + (1-\lambda)[tr[C^T\Gamma CK] - 2tr[C^TL] + d]\} \\ &= (\sigma^2 + \tau^2)[\lambda V(\xi_n) + (1-\lambda)B(\xi_n)].\end{aligned} \tag{8}$$

Since $(\sigma^2 + \tau^2)$ is a constant, a design is $\bar{R}$ -optimal if it minimizes $\lambda V + (1-\lambda)B$. A design is all-variance design if it minimizes V alone, and a design is all-bias design if it minimizes B alone. Let $\xi_{n,\lambda}$ denote the $\bar{R}$ -optimal design associated with λ , Then $\xi_{n,0}$ is the all-bias design, and $\xi_{n,1}$ is the all-variance design.

To compare different designs for model in (2), we define the efficiency of a design by

$$e(\xi_n, \lambda) = \frac{\bar{R}^*(\lambda)}{\bar{R}(\xi_n, \lambda)} \tag{9}$$

where $\bar{R}^*(\lambda)$ is the minimum of $\bar{R}(\xi_n,\lambda)$ over all n -point designs for a given λ .

3 Illustrative Example

In this section we present some numerical results on the all-bias, all-variance and compound optimal designs for several models and also investigate the influence of the heteroscedasticity. Throughout, the design region is the unit cube in R^q , i. e. , $\chi \in [0,1]^q$, a generic point in $[0,1]^q$ is denoted by $\mathbf{x} = (x_1, \cdots, x_q)^T$, and the point $\mathbf{x}_i$ is written as $\mathbf{x}_i = (x_{i1}, \cdots, x_{iq})^T$the response functions are all from the so called Sobolev-Hilbert space (Wahba, 1990, chapter 10), The Sobolev-Hilbert space, S , is a reproducing kernel Hilbert space whose reproducing kernel is

$$Q(\mathbf{x}, \mathbf{t}) = \prod_{l=1}^{q}\left[\frac{4}{3} + \frac{x_l^2 + t_l^2}{2} - \max(x_l, t_l)\right],\ \mathbf{x},\mathbf{t} \in [0,1]^q. \tag{10}$$

The inner product in S is

$$\langle f,g \rangle = \sum_{u \subset \{1,\ldots,q\}} \int_{[0,1]^u} \frac{\partial^u f}{\partial \mathbf{x}_u}\frac{\partial^u g}{\partial \mathbf{x}_u} d\mathbf{x}_u, \tag{11}$$

where the sum over u is taken over all the subsets of the coordinates of χ , $\mathbf{x}_u$ denotes the coordinate projection of χ onto $[0,1]^u$ and $d\mathbf{x}_u = \prod_{l \in u} dx_l$. We also assume that each of the components of the vector f in the model (2) lies in the space S such that the matrix $A = (\langle f_i, f_j \rangle)_{p \times p}$ is non-singular. We define

$$K(\mathbf{x}, \mathbf{t}) = Q(\mathbf{x}, \mathbf{t}) - f^T(\mathbf{x})A^{-1}f(\mathbf{t}). \tag{12}$$

It is known (Wahba,1973) that K in (12) is a reproducing kernel for the subspace of S which is orthogonal complement to the subspace spanned by f. We take this K as the covariance kernel for the bias space H. We assume that the covariance matrix of the errors is that of an AR(1) process:

$$cov[\varepsilon] = \frac{\sigma^2}{1-\rho^2}P_1,\quad P_1(i,j) = \rho^{|i-j|},\ 0 \leqslant |\rho| < 1, \tag{13}$$

where σ^2 is known and the parameter ρ may be estimated from the data through, e. g, the Cochran – Orcuct procedure (Montgomery Peck, 1982, p355), when ρ is unknown. In this paper we also allow for the covariance matrix P_2, that is

$$cov[\varepsilon] = \sigma^2 P_2,\ P_2(i,i) = 1,\ P_2(i,j) = \rho,\ 0 \leqslant |\rho| < 1. \tag{14}$$

We also suppose the parameters σ^2, ρ are known. The reason why we consider the covariance matrix P_2 is that in the view of the robust estimation, the estimator $\hat{\theta} = (X^TX)^{-1}X^TY$, as the estimator $\hat{\theta} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}Y$, is still a best linear estimator in the model (2) with the covariance matrix σ^2P_2. Numerical results indicate that under this situation the designs gettin through the $\bar{R}$ -optimal criterion are more robust against the departures from the assumption of homoscedasticity.

Example In this example we consider the case in which the assumed model is

$$f(\mathbf{x}) = (1, B_1(x_1), \cdots, B_1(x_q))^T,\ B_1(x) = x - \frac{1}{2}.$$

Then the kernel defined in (12) is given by

$$K(\mathbf{x},\ \mathbf{t}) = Q(\mathbf{x},\ \mathbf{t}) - f^T(\mathbf{x})f(\ \mathbf{t}). \tag{15}$$

This means that the response is fitted by a first – degree polynomial and the bias is random with zero mean and covariance kernel $K(\mathbf{x},\ \mathbf{t})$. For the case with $q = 1$ we assume that the regression function is $f_1(x) = (1, B_1(x))^T$. Then the kernel defined in (15) can be expressed as

$$K_1(x,t) = \frac{1}{2}B_2(|x-t|),\ B_2(x) = x^2 - x - \frac{1}{6}. \tag{16}$$

Figure 1 shows some optimal designs $\xi_{n,\lambda}$ in $[0,1]$, and the efficiencies of the all-bias and all-variance designs which are calculated according to the definition in (9) for $\lambda = 0.1, 0.2, \cdots, 0.9$ while $\rho = 0, 0.5, -0.5$, respectively. In this case, we take P_1 as the variance-covariance matrix. Figure 2 (a) shows some optimal designs $\xi_{n,\lambda}$ in $[0,1]$ when $\lambda = 0,0,4,0,8,1$ but $\rho = 0.1, 0.2, \cdots, 0.9$. The covariance matrix that we assume is $P_1 > 0$. Similarly, in Figure 2 (b) we give the designs assuming the covariance matrix is $P_2 > 0$. From figure 1 we observe that the n points of the all-bias design are scattered uniformly in the design domain, and the $\bar{R}$ -optimal design ξ_n tends to the all-bias design as the value of λ tends to zero. However, when the value of λ is getting large in $(0,1)$, the $\bar{R}$ -optimal design tends to the all-variance design. The similar phenomenon (not shown in this paper) could be observed when the covariance matrix was P_2. From figure 2 (a) we find that the n points run to the extreme as λ becomes large. Two special cases are: For $\lambda = 0$ (all-bias design), the n – points are scattered uniformly and is little sensitive to the correlative coefficient ρ; However, in the situation $\lambda = 1$ (all-variance design), the design points are located on the boundaries. For the compound design, the design becomes more sensitive to ρ. But when we choose another covariance matrix P_2, as is shown in (b), the design points will be less sensitive. So we conclude that the $\bar{R}$ -optimal criterion strongly depends on the covariance matrix. Similar computations were done for some other values of n and the negative coefficient ρ, and similar results were also obtained. For the case with $q = 2$ we let

$$f_2(\mathbf{x}) = (1, B_1(x_1), B_1(x_2))^T. \tag{17}$$

The kernel defined in (15) is

$$K_2(\mathbf{x},\ \mathbf{t}) = Q(\mathbf{x},\ \mathbf{t}) - f_2^T(\mathbf{x})f_2(\ \mathbf{t}). \tag{18}$$

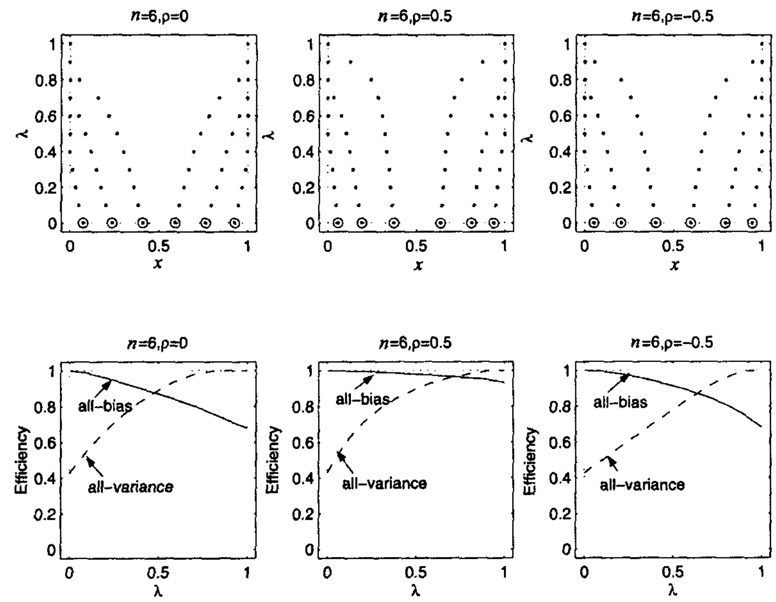
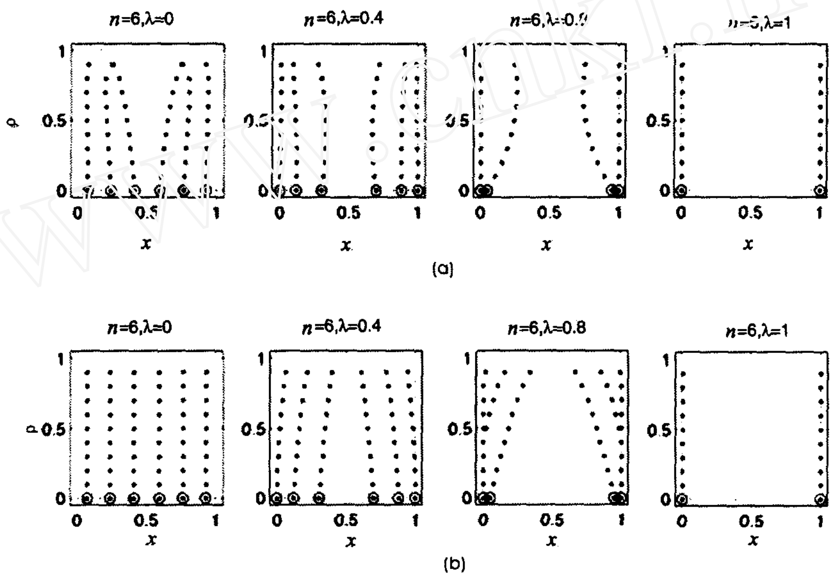


Figure 1 some optimal designs $\xi_{n,\lambda}$ in [0,1], and the efficiencies of the all-bias and all-variance designs when the variance-covariance matrix is P_1



(a) $P_1 > 0$ (b) $P_2 > 0$.

Figure 2 some optimal designs $\xi_{n,\lambda}$ in [0,1] when $\lambda = 0,0,4,0,8,1$ but $\rho = 0.1,0.2,\cdots,0.9$.

where $Q(\mathbf{x}, \mathbf{t})$ is defined by (10) with $q = 2$. In Figure 1, we show some designs that minimizes $\bar{R}$given in (8), where $\rho = 0, 0.5, -0.5$ and $\lambda = 0(\cdot), 0.5(*), 1(\circ)$ in Figure 3. The efficiencies of the all-bias and all-variance designs which are calculated according to the definition in (9) are shown in the same figure. As we find in the case $q = 1$, the all-bias design is a uniform design, and the all-variance design points are all the extreme points. Figure 4 (a) shows the optimal designs obtained according to the criterion we give in (8) under the assumption that the model with variance - covariance P_1 when $\lambda = 0, 0.4, 0.8, 1$, but $\rho = 0(\cdot), 0.5(*), 0.9(\circ)$, respectively. Figure 4 (b) shows the designs with $\lambda = 0, 0.4, 0.8, 1$ and $\rho = 0(\cdot), 0.5(*), 0.9(\circ)$, respectively. The variance – covariance matrix that we take is P_2 . From (a) and (b) we get the conclusion similar to the case with $q = 1$. That is, when $\lambda = 0$, uniform designs are shown; At the same time, when $\lambda = 1$, design points appeared in the extreme points. The designs under the two situations are both robust against from ρ . For compound designs, designs got from the model (2) with the covariance matrix P_1 , are less robust than those with the covariance matrix, P_2 .

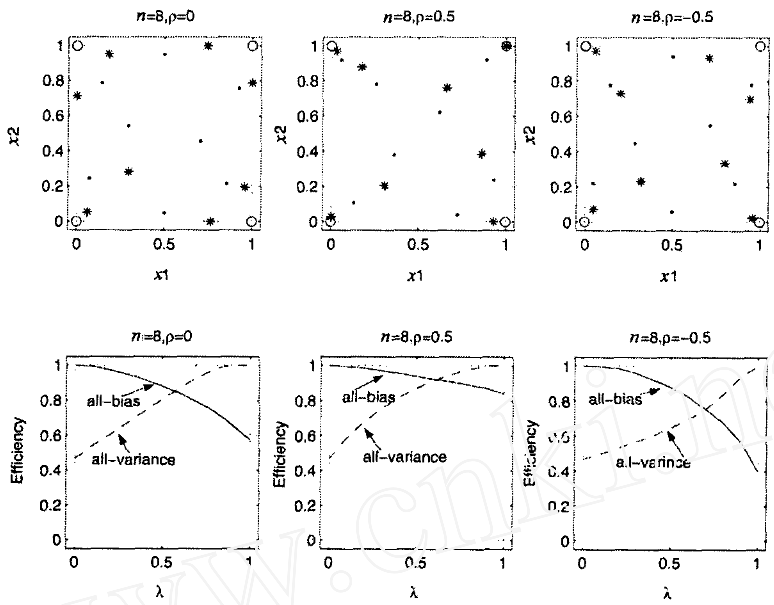
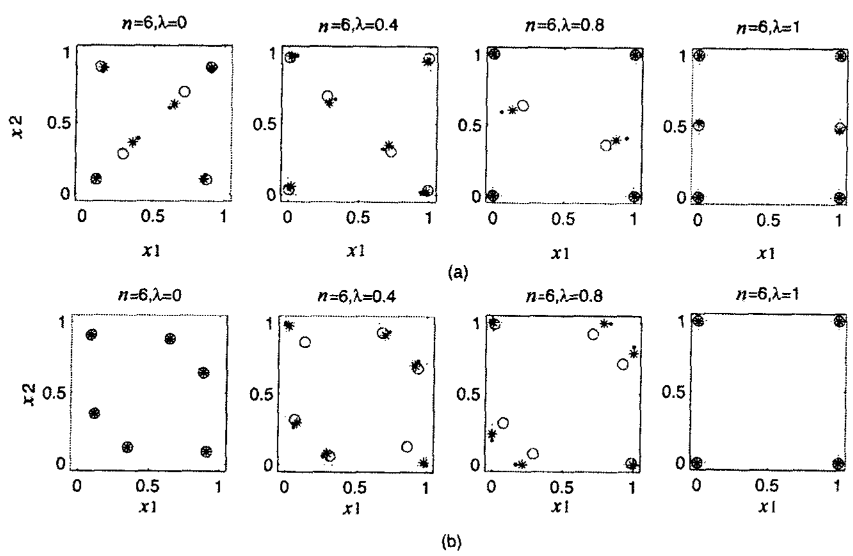


Figure 3　some designs $\xi_{n,\lambda}$ in $[0,1]^2$, and the efficiencies of the all-bias and all-variance designs hen the variance-covariance matrix we take is P_1 .



(a) $P_1 > 0$　　(b) $P_2 > 0$.

Figure 4　the optimal designs $\xi_{n,\lambda}$ in $[0,1]^2$ when $\lambda = 0, 0.4, 0.8, 1$, but $\rho = 0(\cdot), 0.5(*), 0.9(\circ)$.

4 Summary

We have considered the design problem for the response surface model with bias and correlated errors by applying reproducing kernel Hilbert space approach. The criterion for choosing designs is the average loss which can be decomposed into two terms: variance term and bias term. From the numerical examples, we find that when the covariance matrix is known, the average expected loss depends on the quantity λ. For small values of λ in $(0,1)$, the all-bias design is better than the all-variance design, and for the large values of λ the all-variance design is better. For the special structure of the covariance matrix, the designs attained through $\bar{R}$ are much more robust against the departures of assumption of homoscedasticity.

Reference:

[1] JENKIMS G M, CHANMUGAM J. The estimation of slope when the errors are autocorrelated[J]. Journal of the Royal Stiestical Society Series B, 1962,24: 199-214.

[2] KIEFER J. Optimal designs for fitting multiresponse surfaces[J]. Symposium on Multivariate Analysis III, 1973, 287-297. AcademicPress, New York.

[3] LI K C, NOTZ W. Robust designs for nearly linear regression[J]. J Statist Plann Inference 6, 1982(6): 135-151.

[4] PESOTCHINSHY L. Linear estimation for approximately linear models[J]. ANN Statist, 1978(6):1122-1137.

[5] WAHBA G. Improper priors, spline smoothing and the problem of guarding against model errors in regression[J]. J R StatistSoc Ser, 1978,40: 364-372.

[6] WAHBA G. Spline models for observational data[J]. SIAM, Philadephia, 1990.

[7] WEINS D P, ZHOU J L. Minimax desings for approximately linear models with AR(1) errors[J]. The Canadian Journal ofStatistics, 1999, 27: 781-794.

[8] YUE R X, HICKERNELL F J. Robust-designs for fitting linear models with misspecification[J]. Statistica Sinica, 1999 (9):1053-1069.

相依误差近似回归模型的稳健设计

周晓东，岳荣先

(上海师范大学 数理信息学院，上海 200234)

摘　要：利用再生核空间方法研究相依误差近似回归模型的稳健设计问题.假定模型偏差来自概率测度为 P 的函数类 H，同时还假定不同试验所产生的误差不独立.采用广义最小二乘估计作为参数向量的估计，根据平均预测风险最小原理得到了一个稳健设计准则.数值例子表明由该准则得到的设计具有良好的稳健性.

关键词：稳健设计；再生核 Hilbert 空间；相依误差