Chin. Ann. Math.
28B(2), 2007, 177–186
DOI: 10.1007/s11401-005-0351-1

Chinese Annals of Mathematics, Series B

Unbiased Quasi-regression****

Guijun YANG* Lu LIN** Runchu ZHANG***

Abstract Quasi-regression, motivated by the problems arising in the computer experiments, focuses mainly on speeding up evaluation. However, its theoretical properties are unexplored systemically. This paper shows that quasi-regression is unbiased, strong convergent and asymptotic normal for parameter estimations but it is biased for the fitting of curve. Furthermore, a new method called unbiased quasi-regression is proposed. In addition to retaining the above asymptotic behaviors of parameter estimations, unbiased quasi-regression is unbiased for the fitting of curve.

Keywords Computer experiment, Quasi-regression, Unbiasedness, Fitting of curve, Asymptotic normality
2000 MR Subject Classification 62G20, 62J05

1 Introduction

We first outline the quasi-regression proposed by Owen [7] and An and Owen [1]. Consider the problem of approximating an unknown function $f : [0,1]^{p-1} \to R^q$ by function $\widehat{f} : [0,1]^{p-1} \to R^q$. In this paper, we assume $q = 1$, for simplicity.

In [1], the approximating approach begins with an equation

$$f(x) = \beta_0 + \sum_{j=1}^{p-1} z^j(x)\beta_j + \eta(x), \tag{1.1}$$

where β_0 through β_{p-1} are scalar coefficients, $x = (x^1, \cdots, x^{p-1})^T$ is a column vector of design variables, $z^1(x)$ up to $z^{p-1}(x)$ are basis functions satisfying the following conditions

$$\begin{aligned} &\int z^j(x)dx = 0, \quad \int (z^j(x))^2 dx = 1 \quad \text{for } j = 1, \cdots, p-1, \\ &\int z^j(x)z^k(x)dx = 0 \quad \text{for } j \neq k, \end{aligned} \tag{1.2}$$

Manuscript received August 30, 2005. Revised February 28, 2006. Published online March 5, 2007.
*Department of Statistics, Tianjin University of Finance and Economics, Tianjin 300222, China.
E-mail: tjyangguij@yahoo.com.cn
**School of Mathematics and System Sciences, Shandong University, Jinan 250100, China.
E-mail: linlu@sdu.edu.cn
***LPMC and School of Mathematical Sciences, Nankai University, Tianjin 300071, China.
E-mail: zhrch@nankai.edu.cn
****Project supported by the National Natural Science Foundation of China (No. 10571093, No. 10371059), Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20050055038), the Natural Science Foundation of Shandong Province of China (No. 2006A13), the China Postdoctoral Science Foundation (No. 20060390169) and the Tianjin Planning Programs of Philosophy and Social Science of China (No. TJ05-TJ002).

where all integrals are over $[0,1]^{p-1}$. The basis functions may be chosen as sinusoids, wavelet (see [2]), orthogonalized B-spline and so on.

For an unknown function $f(x)$ containing many variables such as 1,000,000 variables, Owen [7] proposed an approach, which begins with a linear equation

$$f(x) = \beta_0 + \sum_{j=1}^{p-1} z^j(x^j)\beta_j + \eta(x). \tag{1.3}$$

In this paper, we only use the equation (1.3) to construct linear approximation by regression and quasi-regression presented below.

Regression-based approaches for computer experiments (see [3, 8]) are defined through the least squares values for $\beta = (\beta_0, \beta_1, \cdots, \beta_{p-1})^T$, i.e.,

$$\beta = \arg\min_{\beta} \int (f(x) - z(x)\beta)^2 dx, \tag{1.4}$$

where $z(x) = (1, z^1(x^1), \cdots, z^{p-1}(x^{p-1}))$. Elementary manipulations give

$$\beta = \Big(\int z(x)^T z(x) dx \Big)^{-1} \int z(x)^T f(x) dx \tag{1.5}$$

$$= \int z(x)^T f(x) dx. \tag{1.6}$$

As a result, the residual function $\eta(x) = f(x) - \beta_0 - \sum_{j=1}^{p-1} z^j(x^j)\beta_j$ satisfies $\int \eta(x)dx = \int \eta(x) z^j(x^j) dx = 0$ for $j = 1$ through $p-1$. In addition, the parameter $\sigma^2 = \int \eta^2(x)dx$ is called the variance of residual $\eta(x)$.

Let $z(x_i) = (1, z^1(x_i^1), \cdots, z^{p-1}(x_i^{p-1}))$ be the row vector of all p basis functions evaluated at the ith input point $x_i = (x_i^1, \cdots, x_i^{p-1})^T$ for $i = 1$ up to n, and Z be the $n \times p$ matrix with ith row $z(x_i)$. Similarly, $Y_i = f(x_i)$, $\varepsilon_i = \eta(x_i)$, and Y denotes the column vector with ith entry Y_i, ε denotes the column vector with ith entry ε_i. In this paper, we assume Y_1 through Y_n are i.i.d. observations. The regression approach is to take an independent Monte Carlo sample $x_i^j \sim U[0,1]$, for $i = 1$ through n and $j = 1$ up to $p-1$, and to estimate the integral in (1.5) by these sample values. This results in

$$\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1, \cdots, \widehat{\beta}_{p-1})^T = (Z^T Z)^{-1} Z^T Y \tag{1.7}$$

and an approximation of $f(x)$ as $\widehat{f}(x) = z(x)\widehat{\beta}$. Quasi-regression exploits the known value $\int z(x)^T z(x) dx = I$, and estimates β in equation (1.6) by an independent Monte Carlo samples, i.e.,

$$\widetilde{\beta} = (\widetilde{\beta}_0, \widetilde{\beta}_1, \cdots, \widetilde{\beta}_{p-1})^T = \frac{1}{n} Z^T Y \tag{1.8}$$

and approximates $f(x)$ by $\widetilde{f}(x) = z(x)\widetilde{\beta}$. Furthermore, quasi-regression has an advantage in evaluation complexity, and the costs for calculating quasi-regression are smaller than those for calculating regression (see [1]). The recent study about quasi-regression focuses mainly on the improvment of its algorithms (see for example [10, 11]).

However, its statistical properties are unexplored systemically. In this paper, we discuss quasi-regression theoretically and numerically in parametric inference and fitting of curve under

general model presented above. Some small sample and large sample properties, including the unbiasedness, strong convergence, asymptotic normality and the property of simulation, are obtained. We also find some defects of quasi-regression arising in numerical simulation, and give a theoretic reason why simply adding basis functions is of no help to the fitting of curve. Then we propose a new method, called unbiased quasi-regression, which is superior to quasi-regression in the sense that the method is unbiased not only for parameter estimations but also for the fitting of curve and retains the above asymptotic behaviors of parameter estimations. An example is investigated numerically to illustrate the theoretical conclusions.

2 Properties of Parameter Estimations

To study the statistical properties of quasi-regression, we firstly calculate the moments of the estimators for scalar coefficients in (1.1). We have the following theorem.

Theorem 2.1 (i) *If* $E(f^2(x))$ *exists, then*

$$E(\widetilde{\beta}_j) = \beta_j \quad \textit{for } j \geq 0$$

and

$$\mathrm{Var}(\widetilde{\beta}_0) = \frac{1}{n}(E(f^2(x)) - \beta_0^2).$$

(ii) *If* $E(f^4(x))$ *and* $E((z^j(x^j))^4)$ *exist, then*

$$\mathrm{Var}(\widetilde{\beta}_j) = \frac{1}{n}(E((z^j(z^j))^2 f^2(x)) - \beta_j^2) \quad \textit{for } j \geq 1.$$

This theorem can be proved by definition (1.3) and condition (1.2), and its proof is omitted here. From Theorem 2.1, we can obtain the convergency in mean square error. The following corollary gives the details.

Corollary 2.1 *Under the conditions of Theorem* 2.1, $E(\widetilde{\beta}_j - \beta_j)^2 \to 0$ *for* $j \geq 0$ *as* $n \to \infty$.

Since $\widetilde{\beta}$ is indeed the sum of independent and identically distributed random variables, from Theorem 2.1, the Kolmogorov strong law of large numbers and the center limit theorem, we can get the following corollaries concerning the strong convergency and asymptotic normality.

Corollary 2.2 *Under the conditions of Theorem* 2.1, $\widetilde{\beta} \xrightarrow{a.s.} \beta$ *as* $n \to \infty$.

Corollary 2.3 *Under the conditions of Theorem* 2.1, $\sqrt{n}\,(\widetilde{\beta}_0 - \beta_0) \xrightarrow{\mathcal{D}} N(0, E(f^2(x)) - \beta_0^2)$ *and* $\sqrt{n}\,(\widetilde{\beta}_j - \beta_j) \xrightarrow{\mathcal{D}} N(0, E((z^j(z^j))^2 f^2(x)) - \beta_j^2)$ *for* $j \geq 1$.

Theorem 2.1 shows the quasi-regression estimator $\widetilde{\beta}$ is unbiased. However, from [7, Proposition 5.1], the regression estimator $\widehat{\beta}_j$ is biased and its bias is

$$E(\widehat{\beta}_j) - \beta_j = -\frac{1}{n}E(z^j(x^j)\eta(x)S^2(x)) = -\frac{p}{n}E(z^j(x^j)\eta(x)) + O\Big(\frac{p^{1/2}}{n}\Big) = O\Big(\frac{p^{1/2}}{n}\Big),$$

where $S^2(x) = \sum\limits_{j=1}^{p-1}(z^j(x^j))^2$. On the other hand, we can also compare the variance of $\widetilde{\beta}_j$ with the variance of $\widehat{\beta}_j$. Theorem 2.1 shows that the asymptotic order of variance of quasi-regression

estimator $\widetilde{\beta}_j$ is $O(\frac{1}{n})$. From [7, Lemma 5.1], we have

$$E(\widehat{\beta}_j^2) = \beta_j^2 - \frac{2\beta_j}{n}E(S^2(x)(z^j(x^j))^2\eta^2(x)) + \frac{1}{n}E((z^j(x^j))^2\eta^2(x)) + o\Big(\frac{1}{n}\Big).$$

In addition, for any function $h(x)$, if $\int h^4(x)dx < \infty$, then

$$E(S^2(x)h^2(x)) = (p-1)E(h^2(x)) + O(p^{1/2}) \tag{2.1}$$

(cf. [7, p. 6]). Therefore

$$\begin{aligned} E(\widehat{\beta}_j^2) &= \beta_j^2 - \frac{2\beta_j(p-1)}{n}E((z^j(x^j))^2\eta^2(x)) + \frac{1}{n}E((z^j(x^j))^2\eta^2(x)) + O\Big(\frac{p^{1/2}}{n}\Big) \\ &= \beta_j^2 + O\Big(\frac{p}{n}\Big). \end{aligned}$$

Then for $j \neq 0$,

$$\mathrm{Var}(\widehat{\beta}_j) = E(\widehat{\beta}_j^2) - (E(\widehat{\beta}_j))^2 = O\Big(\frac{p}{n}\Big).$$

From the results above, we conclude that quasi-regression is better than regression in the sense of unbiasedness and asymptotic variance of estimators for scalar coefficients.

As for the estimator of σ^2, Owen [7] provided the linear variation $\sigma_L^2 = \int(f_L(x) - \beta_0)^2dx = \sum_{j=1}^{p-1}\beta_j^2$ and the nonlinear variation $\sigma_{NL}^2 = \int(f(x) - f_L(x))^2dx = \int(\eta(x))^2dx$ in $f(x)$, where $f_L(x) = \beta_0 + \sum_{j=1}^{p-1} z^j(x^j)\beta_j$. It can be verified that the nonlinear variation is equal to σ^2. Owen [7] took $\sum_{j=1}^{p-1}\widetilde{\beta}_j^2$ as estimator of the linear variation σ_L^2. And the nonlinear variation σ_{NL}^2, i.e., variance σ^2, is estimated by subtracting $\sum_{j=1}^{p-1}\widetilde{\beta}_j^2$ from the total variation $\frac{1}{n}\sum_{i=1}^{n}(f(x_i) - \overline{f})^2$, where $\overline{f} = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$. The estimator of σ^2 is

$$\widetilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - \overline{f})^2 - \sum_{j=1}^{p-1}\widetilde{\beta}_j^2.$$

From [7, Proposition 4.1], we have

$$E\Big(\sum_{j=1}^{p-1}\widetilde{\beta}_j^2\Big) = \frac{n-1}{n}\sum_{j=1}^{p-1}\beta_j^2 + \frac{1}{n}E(S^2(x)f^2(x)).$$

On the other hand, it can be verified that

$$E\Big(\frac{1}{n}\sum_{i=1}^{n}(f(x_i) - \overline{f}\,)^2\Big) = \frac{(n-1)(E(f^2(x)) - \beta_0^2)}{n}$$

and

$$\sigma^2 = \sigma_{NL}^2 = \int(f(x) - f_L(x))^2dx = E(f^2(x)) - \beta_0^2 - \sum_{j=1}^{p-1}\beta_j^2.$$

Therefore, we obtain the following theorem.

Theorem 2.2 *If* $E(z^j(x^j))^4$ *and* $E(f(x))^4$ *exist, then* $E(\widetilde{\sigma}^2) = \frac{n-1}{n}\sigma^2 - \frac{1}{n}E(S^2(x)f^2(x))$. *Consequently,* $E(\widetilde{\sigma}^2) = \sigma^2 - \frac{p}{n}E(f^2(x)) + O(\frac{p^{1/2}}{n}) = \sigma^2 + O(\frac{p}{n})$.

Analogously, using (1.7), we get another estimator of σ^2, denoted by $\widehat{\sigma}^2$, which is

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - \overline{f})^2 - \sum_{j=1}^{p-1}\widehat{\beta}_j^2. \tag{2.2}$$

Owen [7] showed that if $E(z^j(x^j))^4$ and $E(\eta(x))^4$ exist, then

$$E\Big(\sum_{j=1}^{p-1}\widehat{\beta}_j^2\Big) = \sum_{j=1}^{p-1}\beta_j^2 + \frac{1}{n}E(S^2(x)\eta^2(x)) + O(n^{-2}).$$

Using (2.2) and the above method, we get

$$E(\widehat{\sigma}^2) = \frac{(n-1)\sigma^2}{n} - \frac{1}{n}\sum_{j=1}^{p-1}\beta_j^2 - \frac{1}{n}E(S^2(x)\eta^2(x)) + O(n^{-2}).$$

From (2.1), we have

$$E(\widehat{\sigma}^2) = \sigma^2 - \frac{p}{n}E(\eta^2(x)) + O\Big(\frac{p^{1/2}}{n}\Big).$$

It can be verified that $E(f^2(x)) > E(\eta^2(x))$. Consequently, the bias of $\widehat{\sigma}^2$ is smaller than of $\widetilde{\sigma}^2$. However, both $\widetilde{\sigma}^2$ and $\widehat{\sigma}^2$ have the same asymptotic orders.

3 Unbiased Quasi-regression

We have to realize that quasi-regression has some imperfections. Denote the residual of quasi-regression by $\widetilde{\eta}(x_i)$. Then

$$\widetilde{\eta}(x_i) = f(x_i) - \widetilde{f}(x_i), \quad i = 1, \cdots, n,$$

where fitted value $\widetilde{f}(x_i) = \widetilde{\beta}_0 + \sum\limits_{j=1}^{p-1} z^j(x_i^j)\widetilde{\beta}_j$. From (2.1), we have

$$E(\widetilde{\eta}(x_i)) = -\frac{1}{n}E\Big(\sum_{j=1}^{p-1}z^j(x_i^j)\sum_{s=1}^{n}z^j(x_s^j)f(x_s)\Big) = -\frac{p-1}{n}\beta_0 + O\Big(\frac{p^{1/2}}{n}\Big). \tag{3.1}$$

This result implies that $|E(\widetilde{\eta}(x_i))|$ is very large, if $\beta_0 \neq 0$ and p is very large. The residuals will increase with number of basis functions (or design variables). For the model containing large numbers of variables, this bias is more serious. This gives a theoretical explanation for the problem appearing in numerical simulation. Some examples of numerical simulation in [1] showed that simply adding basis functions does not improve fitting of curve. Now, we provide an alternative fitting model of (1.3), where fitted value is

$$\underline{f}(x_i) = \underline{\beta_0} + \sum_{j=1}^{p-1}z^j(x_i^j)\widetilde{\beta}_j, \tag{3.2}$$

where $\widetilde{\beta}_1$ through $\widetilde{\beta}_{p-1}$ are provided in (1.8) and $\underline{\beta_0} = \frac{1}{n}\sum\limits_{i=1}^{n}\Big(1-\frac{1}{n}\sum\limits_{j=1}^{p-1}(z^j(x_i^j))^2\Big)f(x_i)$. In the new model, the residual, denoted by $\underline{\eta}(x_i) = f(x_i) - \underline{f}(x_i)$, has the following properties.

Theorem 3.1 (i) *If* $E(f^2(x))$ *exists, then* $E(\underline{\eta}(x_i)) = 0$ *and* $E(\widetilde{\eta}(x_i)) = -\frac{p-1}{n}\beta_0 + O\big(\frac{p^{1/2}}{n}\big)$.

(ii) *If* $E(f^6(x))$ *and* $E((z^j(x^j))^4)$ *exist, then* $\mathrm{Var}(\underline{\eta}(x_i)) = \mathrm{Var}(\eta(x)) + O\big(\frac{p^2}{n}\big)$ *and* $\mathrm{Var}(\widetilde{\eta}(x_i)) - \mathrm{Var}(\underline{\eta}(x_i)) = \frac{p^2\beta_0^2}{n^2} + O(\frac{p}{n^2}) + O(\frac{p^2}{n^3})$.

Proof (i) can be immediately verified by (3.1). For (ii), let $\widetilde{\beta}_j(i) = \frac{1}{n}\sum\limits_{1\le s\le n, s\ne i} z^j(x_s^j)f(x_s)$. From Theorem 2.1 and equation (2.1), we have

$$\mathrm{Var}(\underline{\eta}(x_i)) = E(\underline{\eta}(x_i))^2 = E(f^2(x)) - \beta_0^2 - \sum_{k=1}^{p-1}\beta_k^2 + O\Big(\frac{p^2}{n}\Big) = \mathrm{Var}(\eta(x)) + O\Big(\frac{p^2}{n}\Big).$$

Similarly, from (3.1) and the method used above, it follows that

$$\mathrm{Var}(\widetilde{\eta}(x_i)) - \mathrm{Var}(\underline{\eta}(x_i)) = E(\widetilde{\eta}(x_i))^2 - (E(\widetilde{\eta}(x_i)))^2 - E(\underline{\eta}(x_i))^2 = \frac{p^2\beta_0^2}{n^2} + O\Big(\frac{p}{n^2}\Big) + O\Big(\frac{p^2}{n^3}\Big).$$

The proof is completed.

Table 3.1 Quasi-regression and unbiased quasi-regression

Estimation Complexity	Time	Space
Regression	$O(np^2+p^3)$	$O(p^2)$
Quasi-regression	$O(np)$	$O(p)$
Unbiased quasi-regression	$O(np+p)$	$O(p)$
β_0	Expectation	Variance
Quasi-regression	β_0	$O(\frac{1}{n})$
Unbiased quasi-regression	$\beta_0 + O(\frac{p}{n})$	$O(\frac{p^2}{n})$
$\beta_i, i = 1,\cdots,p-1$	Expectation	Variance
Quasi-regression	β_i	$O(\frac{1}{n})$
Unbiased quasi-regression	β_i	$O(\frac{1}{n})$
Residual	Expectation	Variance
Quasi-regression	$-\frac{p-1}{n}\beta_0 + O\big(\frac{p^{1/2}}{n}\big)$	$E(f^2(x)) - \beta_0^2 - \sum\limits_{k=1}^{p-1}\beta_k^2 + O\big(\frac{p^2}{n}\big)$
Unbiased quasi-regression	0	$E(f^2(x)) - \beta_0^2 - \sum\limits_{k=1}^{p-1}\beta_k^2 + O\big(\frac{p^2}{n}\big)$

From Theorem 3.1, the residual $\underline{\eta}(x_i)$ satisfies $E(\underline{\eta}(x_i)) = 0$, so the new model is called unbiased quasi-regression. For n data points and p basis functions, we provide the result of comparing unbiased quasi-regression with quasi-regression in Table 3.1. In Table 3.1, unbiased quasi-regression has the same evaluation complexity as quasi-regression. In (3.2), all estimators of β_j for $j \ge 1$ do not change except that of β_0, which changes from $\widetilde{\beta}_0 = \frac{1}{n}\sum\limits_{i=1}^{n} f(x_i)$ to $\underline{\beta_0} = \frac{1}{n}\sum\limits_{i=1}^{n}\Big(1-\frac{1}{n}\sum\limits_{j=1}^{p-1}(z^j(x_i^j))^2\Big)f(x_i)$. The only loss in parameter estimation caused by this

changes is that $\underline{\beta_0}$ is biased and the bias is $O(\frac{p}{n})$. But, we get better residual $\underline{\eta}(x)$, because $\underline{\eta}(x)$ has mean zero and its variance is smaller than that of $\widetilde{\eta}(x_i)$ when $\beta_0 \neq 0$. Although $\mathrm{Var}(\underline{\eta}(x_i))$ will increase at rate $O(\frac{p^2}{n})$ with p, this loss can not be avoided under general condition and common statistical methods, for example, quasi-regression, behave like this. Hence for fitting of curve, $\underline{f}(x_i)$ is superior to $\widetilde{f}(x_i)$ in statistical effectiveness.

Like the regression, we can use residual sum of squares to estimate the variance of residual. Note that the residual of quasi-regression is bad, while the residual of unbiased quasi-regression has some better properties. Denote the mean square residual of unbiased quasi-regression by $\underline{R}^2$, and $\underline{R}^2 = \frac{1}{n}\sum\limits_{i=1}^{n}\underline{\eta}^2(x_i)$. Theorem 3.1 shows that $\underline{R}^2$ is an asymptotically unbiased estimator of σ^2. The following corollary gives the details.

Corollary 3.1 *If* $E(f^6(x))$ *and* $E((z^j(x^j))^4)$ *exist, then* $E(\underline{R}^2) = \sigma^2 + O(\frac{p^2}{n})$.

Corollary 3.1 and Theorem 2.2 show that $\widetilde{\sigma}^2$ is better than $\underline{R}^2$ in estimation effectiveness. We provide a new estimator of σ^2, denoted by $\underline{\sigma}^2$, which is

$$\underline{\sigma}^2 = \frac{n}{n-1}\Big(\widetilde{\sigma}^2 + \frac{1}{n^2}\sum_{i=1}^{n} S^2(x_i)f^2(x_i)\Big).$$

From Theorem 2.2, we have the following corollary.

Corollary 3.2 *If* $E(z^j(x^j))^4$ *and* $E(f(x))^4$ *exist, then* $E(\underline{\sigma}^2) = \sigma^2$.

From Corollary 3.2, the estimator $\underline{\sigma}^2$ is unbiased, and requires extra computation on the order of np to calculate the sample average of $S^2(x)f^2(x)$. This evaluation complexity is the same as that of $\widetilde{\sigma}^2$. As a result, the costs for calculating $\underline{\sigma}^2$ is smaller than those for calculating estimator $\widehat{\sigma}^2$, i.e., new estimator $\underline{\sigma}^2$ has more favorable estimation complexity than $\widehat{\sigma}^2$.

4 Numerical Results

To fix ideas, we consider the borehole function investigated in [1, 6]. The function is defined as

$$f(T_u, H_u, H_l, r, r_\omega, L, K_\omega, T_l) = \frac{2\pi T_u(H_u - H_l)}{\log(\frac{r}{r_\omega})\Big(1 + \frac{2LT_u}{\log(\frac{r}{r_\omega})r_\omega^2 K_\omega} + \frac{T_u}{T_l}\Big)}. \tag{4.1}$$

This is a model for the flow rate of water from an upper to a lower aquifer. The inputs r and r_ω are radii of the borehole and surrounding basin respectively, T_u and T_l are transmissivities of aquifers, H_u and H_l are their potentiometric heads, L is the length of borehole and K_ω is a conductivity. The ranges of 8 input variables respectively are

$$r_\omega \in [0.05, 0.15]m,\ \ r \in [100, 50000]m,\ \ T_u \in [63070, 115600]\frac{m^3}{yr},\ \ T_l \in [63.1, 116]\frac{m^3}{yr},$$
$$H_u \in [990, 1110]m,\ H_l \in [700, 820]m,\ \ L \in [1120, 1680]m,\ \ K_\omega \in [9855, 12045]\frac{m}{yr}.$$

For simplicity, we assume the variables have the same significance. In the example, the sample size is $n = 100$ and the univariate orthogonal basis functions chosen to work with are

$$z^0(x^0) = 1, \quad z^j(x^j) = \frac{1}{\sqrt{6}}\sum_{k=1}^{6}\phi_k(x^j) \quad \text{for } j = 1, \cdots, 8, \tag{4.2}$$

where $\phi_0(x) = 1$ up to $\phi_6(x)$ are the first seven univariate orthogonal polynomials. We select $x_i^j \sim U[0,1]$, hence

$$
\begin{aligned}
r_\omega &= 0.05 + x^1(0.15 - 0.05) \sim U[0.05, 0.15], \quad r = 100 + x^2(50000 - 100) \sim U[100, 50000],\\
T_u &= 63070 + x^3(115600 - 63070) \sim U[63070, 115600],\\
T_l &= 63.1 + x^4(116 - 63.1) \sim U[63.1, 116],\\
H_u &= 990 + x^5(1110 - 990) \sim U[990, 1110], \quad H_l = 700 + x^6(820 - 700) \sim U[700, 820],\\
L &= 1120 + x^7(1680 - 1120) \sim U[1120, 1680],\\
K_\omega &= 9855 + x^8(12045 - 9855) \sim U[9855, 12045].
\end{aligned}
$$

Figure 4.1 shows the relative residuals of quasi-regression defined as $\widetilde{\eta}_r(x_i) = \frac{\widetilde{\eta}(x_i)}{\overline{f}}$ and $\overline{f} = \frac{1}{n}\sum\limits_{i=1}^{n} f(x_i)$. Figure 4.2 shows the relative residuals of unbiased quasi-regression defined as $\underline{\eta_r}(x_i) = \frac{\eta(x_i)}{\overline{f}}$.

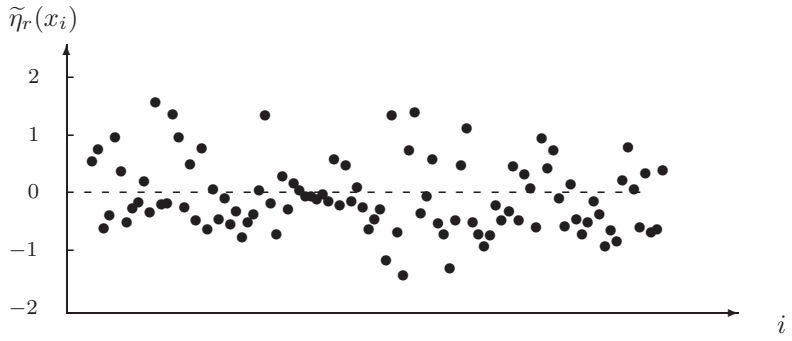

Figure 4.1 The relative residual of quasi-regression

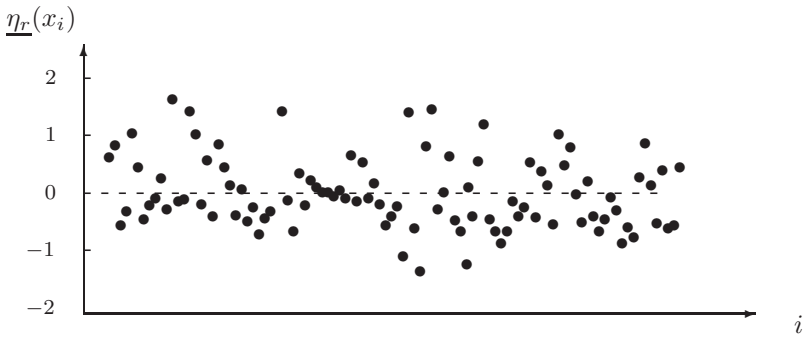

Figure 4.2 The relative residual of unbiased quasi-regression

From the above figures, the residual points of unbiased quasi-regression are evenly distributed on both sides of the center line, while the residual points of quasi-regression are biased towards below the center line. According to numerical simulation, we obtain that the original mean relative residual of squares is $\widetilde{R}_r^2 = \frac{1}{n}\sum\limits_{i=1}^{n} \widetilde{\eta}_r^2(x_i) = 0.3915$, the new mean relative residual of squares is $\underline{R}_r^2 = \frac{1}{n}\sum\limits_{i=1}^{n} \underline{\eta_r}^2(x_i) = 0.3841$, the original mean residual of squares is

$\widetilde{R}^2 = \frac{1}{n}\sum_{i=1}^{n}\widetilde{\eta}^2(x_i) = 1850.3521$, the new mean residual of squares is $\underline{R}^2 = \frac{1}{n}\sum_{i=1}^{n}\underline{\eta}^2(x_i) = 1815.3527$, the original residual sum is $\widetilde{R} = \sum_{i=1}^{n}\widetilde{\eta}(x_i) = -6.1267$, and the new residual sum is $\underline{R} = \sum_{i=1}^{n}\underline{\eta}(x_i) = -1.5927$. These numerical results indicate that unbiased quasi-regression is improved greatly in the bias and small in the relative residual sum of squares and residual sum of squares. Hence unbiased quasi-regression is better than quasi-regression in the fitting of curve.

According to the numerical results, function values $f(x_i)$ range on $[13.0015, 187.2624]$, and if $f(x_i)$ are very large or very small, for example $f(x_i) > 100$ or $f(x_i) < 50$, the residuals are very large; otherwise, the residuals are very small. This behaviour is similar to that of regression.

In this numerical simulation, the residual sums of squares are very large. The main causes are that, first, in the computer experiment, the variables scatter widely, resulting in a low efficiency in fitting of curve; second, the construction of quasi-regression is too simple to fit a curve. In the fitted curve, all basis functions are chosen to be the polynomials of degree 6 while the degrees of some variables in $f(x)$ are perfectly possible higher than 6 and others are lower than 6. However, looking at equation (4.1) does not easily let us know exactly the degree of each variable. If we choose the basis functions of higher degree to work with, data can be well fitted but the resulting models lack prediction power, a phenomenon known as overfitting, because overfitting leads to a bigger variance in estimation and prediction (see [9]). We also choose the basis functions of lower degree to work with, for example, let

$$z^0(x^0) = 1, \quad z^j(x^j) = \phi_3(x^j) \quad \text{for } j = 1, \cdots, 8,$$

and $n = 100$, resulting in

$$\widetilde{R}_r^2 = 0.3878, \ \ \underline{R}_r^2 = 0.3700, \ \ \widetilde{R}^2 = 1813.4703, \ \ \underline{R}^2 = 1749.0063, \ \ \widetilde{R} = -8.5795, \ \ \underline{R} = -3.0239.$$

In this case, the biases are relatively large while the residuals are relatively small.

On the other hand, the similar results are seen when the experiment is finished at other values of n. The efficiency to fit curve can increase with n, but the increase is limited, because the variance σ^2 of residual, i.e., the nonlinear variation in $f(x)$, can not be reduced. For example, let $n = 1000$ and basis functions be in (4.2), then the numerical results are

$$\widetilde{R}_r^2 = 0.2932, \ \ \underline{R}_r^2 = 0.2931, \ \ \widetilde{R}^2 = 1701.5741, \ \ \underline{R}^2 = 1701.2535, \ \ \widetilde{R} = -0.5868, \ \ \underline{R} = -0.1503.$$

From this numerical example, we get another conclusion, i.e., with n increasing, the difference between the statistical properties of quasi-regression and unbiased quasi-regression is decreasing. So, only if $\frac{p}{n}$ is not very small, unbiased quasi-regression is obviously superior to quasi-regression, which just illustrates Theorem 3.1.

5 Discussion

The above results are based on the assumption that Y_1 up to Y_n are i.i.d. observations. If Y_1 through Y_n are dependent observations, we use blocks of observations as discussed by Dimitris and Joseph [4] to construct blockwise quasi-regression.

Let M and L be integers depending only on n, satisfying $M = O(n^{1-\tau}),\ 0 < \tau < 1,\ \frac{M}{L} \to c,\ 0 < c \leq 1,$ as $n \to \infty$. Denote

$$B_i = (Y_{(i-1)L+1}, \cdots, Y_{(i-1)L+M})', \quad i = 1, \cdots, Q, \quad Q = \left[\frac{n-M}{L}\right] + 1,$$

where $[x]$ stands for the integer party of x, B_i is a block of observations, M is the window-width, L is the separation between the block starting points, $L = O(n^{1-\tau})$ and $Q = O(n^{\tau})$. For simplicity, the functions $T_{i,M,L}(B_i)$ are chosen as $T^j_{i,M,L}(B_i) = \frac{1}{M}\sum\limits_{s=1}^{M} z^j(x^j_{(i-1)L+s})Y_{(i-1)L+s}, j = 0, 1, \cdots, p-1$, where $z^0(x) \equiv 1$. Thus, the blockwise quasi-regression estimator of β_j is

$$\overline{\beta}_j = \frac{1}{Q}\sum_{i=1}^{Q} T^j_{i,M,L}(B_i), \quad j = 0, 1, \cdots, p-1.$$

For the weakly dependent stationary observations Y_1 up to Y_n, we can prove by the method used in [5] that blockwise quasi-regression has the superiority in parametric statistical inference.

Acknowledgements The authors would like to thank the referees for their helpful comments which lead to the improvement of this article.

References

[1] An, J. and Owen, A. B., Quas-regression, *J. of Complexity*, **17**, 2001, 588–607.

[2] Chui, C. K., An Introduction to Wavelets, Academic Press, Boston, 1992.

[3] Currin, C., Mitchell, T., Morris, M. and Ylvisaker, D., Bayesion prediction of deterministic functions, with applications to the design and analysis of computer experiments, *J. Amer. Statist. Assoc.*, **86**, 1993, 953–963.

[4] Dimitris, N. and Joseph, P., A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation, *Ann. Statist.*, **20**, 1992, 1985–2007.

[5] Lin, L. and Zhang, R. C., Blockwise empirical euclidean likelihood for weakly dependent processes, *Statist. Probab. Lett.*, **53**, 2001, 143–152.

[6] Morris, M. D., Mitchel T. J. and Ylvisaker D., Bayesian design and analysis of Computer experiments: use of derivative in surface prediction, *Technometrics*, **35**, 1993, 243–255.

[7] Owen, A. B., Assessing linearity in high dimensions, *Ann. Statist.*, **28**, 2000, 1–19.

[8] Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, M. H., Design and analysis of computer experiments, *Statist. Sci.*, **4**, 1989, 409–435.

[9] Weisbery, S., Applied Linear Regression, Second edition, John Wiley and Sons, New York, 1985.

[10] Yang, G. J. and Lin, L., Selection of basis functions and improvement of algorithms for quasi-regression (in Chinese), *Chinese Acta Scientiarum Naturalium Universitatis Nankaiensis*, **36**, 2003, 44–49.

[11] Yang, G. J. and Zhang, R. C., Scalar coefficient estimators of quasi-regression without correlativity, *Chinese Acta Scientiarum Naturalium Universitatis Nankaiensis*, **37**, 2004, 51–57.