Research Article

Weighted Semiparameter Model and Its Application

Zhengqing Fu,[1] Guolin Liu,[2] Ke Zhao,[2] and Hua Guo[1]

[1] *College of Mathematics and System Sciences, SDUST, Qingdao, Shandong 266590, China*
[2] *Geomatics College, SDUST, Qingdao, Shandong 266590, China*

Correspondence should be addressed to Guolin Liu; gliu@sdust.edu.cn

A weighted semiparameter estimate model is proposed. The parameter components and nonparameter components are weighted. The weights are determined by the characters of different data. Simulation data and real GPS data are both processed by the new model and least square estimate, ridge estimate, and semiparameter estimate. The main research method is to combine qualitative analysis and quantitative analysis. The deviation between estimated values and the true value and the estimated residuals fluctuation of different methods are used for qualitative analysis. The mean square error is used for quantitative analysis. The results of experiment show that the model has the smallest residual error and the minimum mean square error. The weighted semiparameter estimate model has effectiveness and high precision.

1. Introduction

With the development of surveying and mapping technology, there are more and more nonlinear problems [1, 2]. When we solve the nonlinear problems, linearization is the main method. Owing to the nonlinear characteristics of model, the linearization methods maybe cause large model error [3, 4].

Parameter estimation is primarily solving the unknown parameters according to certain model and criterion [5]. At the same time, nonparameter estimations are flexible in reducing modeling biases [6]. There are both parameter components and nonparameter components in semiparameter estimate model [7, 8]. Semiparameter model is more efficient. It can be superior to the limitation of nonparametric model and parametric model. Semiparameter has been applied in many fields. Its properties has been studied extensively.

We propose a weighted semiparameter estimate model. The parameter components and nonparameter components are weighted against the characters of different data. The minimal error of mean square can be got by adjusting the weights of parameter components and nonparameter components. The most appropriate weights correspond to the minimal mean square error. Then the weights are confirmed through this method. Considering the characters of different data, the weighted semiparameter estimate model improves

the equally weighted semiparameter model. The weighted semiparameter estimate model is contrasted with LS, ridge estimate, and semiparameter estimate. The simulated data and leveling network date are used to check the model. We take the deviation between estimated values and the true value, the estimated residuals fluctuation, and the mean square error as the index. The results of the experiments show the feasibility and effectiveness of model.

2. Four Models

2.1. Nonlinear Least Squares Model. Nonlinear model is defined as follows:

$$L = f(x) + \Delta, \tag{1}$$

where $L = (L_1, L_2, \ldots, L_n)^T$ is $n \times 1$ vector; f is nonlinear function; $X = (X_1, X_2, \ldots, X_t)^T$ is $t \times 1$ vector to be estimated; $\Delta = (\Delta_1, \Delta_2, \ldots, \Delta_n)^T$ is $n \times 1$ the error of observing vector. The first order Taylor series expansion of (1) is

$$L = f(X_0) + \left. \frac{\partial f(X)}{\partial X} \right|_{X_0} (X - X_0) + \Delta. \tag{2}$$

We are setting $(\partial f(X)/\partial X)|_{X_0} = B$. Least square principle needs to be satisfied; that is,

$$\Delta^T P \Delta = (B(X - X_0) + f(X_0) - L)^T \\ \times P(B(X - X_0) + f(X_0) - L) = \min. \tag{3}$$

The least square solution of (3) is

$$\widehat{X}_p = X_0 + (B^T P B)^{-1} B^T P (L - f(X_0)). \tag{4}$$

The corresponding residual error is

$$V_p = (B^T P B)^{-1} B^T P (L - f(X_0)) + f(X_0) - L. \tag{5}$$

The least square solution of equation has relationship with intrinsic curvature and parameter effect curvature [9]. Only in this case, nonlinear degree is low and the estimate value is close to the true value. Otherwise, error is big.

2.2. Ridge Estimate. Ridge estimate is one of the most common biased methods when we process ill-conditioning problem [10, 11]. It can reduce the mean squared error.

Ridge estimate model is defined as follows:

$$\widehat{X} = (N + kI)^{-1} B^T P L, \tag{6}$$

where $0 \le k < \infty$ is ridge parameters and $N = B^T P B$. When $k = 0$, ridge estimate is LS estimate. The determination of ridge parameters is the most important thing in ridge estimate. The main methods to determined ridge parameters are ridge track method, L curve method, and generalized cross-validation (GCV) method [12]. Ridge track method is chosen in this paper.

2.3. Semiparameter Model. Semiparameter model is defined as follows:

$$L = BX + S + \Delta, \tag{7}$$

where $L = (L_1, L_2, \ldots, L_n)^T$ is $n \times 1$ vector; $X = (X_1, X_2, \ldots, X_t)^T$ is $t \times 1$ vector to be estimated; $S = (s(t_1), s(t_2), \ldots, s(t_n))^T$, $s(t_i)$ $0 < t_i < 1$; $\Delta = (\Delta_1, \Delta_2, \ldots, \Delta_n)^T$ is $n \times 1$ independent; and identically distributed random error, $E(\Delta_i) = 0$, $E(\Delta_i^2) < \infty$, $0 \le t_1 < \cdots t_n \le 1$, t_i, and Δ_i are independent.

Linear model is used to estimate the parameter component. Weight function is used to estimate the nonparameter component. On the selected point, weight function $W_i(t_k)$ is defined as follows [13]:

$$w_i(t_k) = \frac{[K(t_k - t_i)/h^{-1}]}{\sum_{j=1}^{n}[K(t_k - t_j)/h^{-1}]}, \quad i, j, k = 1, 2, \ldots, n, \tag{8}$$

where $K(\cdot)$ is kernel function and h is the corresponding window width.

Least square cross-validation method (LSCV) is used to determine the best window width [14]. $\widehat{f}(x)$ is the kernel estimate of probability density function (PDF) $f(x)$. LSCV can minimize the integrated squared error (ISE) of $\widehat{f}(x)$, where

$$\text{ISE}(\widehat{f}(x)) = \int [\widehat{f}(x) - f(x)]^2 dx \\ = \int \widehat{f}^2(x) dx + \int f^2(x) dx - 2 \int \widehat{f}(x) f(x) dx. \tag{9}$$

It is assumed that x is known. The nonparametric kernel estimation of $s(t_k)$ is

$$\widehat{s}(t_k, X) = \sum_{i=1}^{n} W_i(t_k)(L_i - B_i X). \tag{10}$$

The residual error of estimate value is

$$V_k = B_k X + \widehat{s}(t_k, X) - L_k. \tag{11}$$

The vector form of (11) is $V = (I - W)(BX - L)$.

The least square criterion is

$$V^T P V = (BX - L)^T (I - W)^T P (I - W)(BX - L) = \min. \tag{12}$$

The normal equation of (12) is

$$B^T (I - W)^T P (I - W) B X = B^T (I - W)^T P (I - W) L. \tag{13}$$

If B is nonsingular, the semiparametric least squares estimation of X is

$$\widehat{X}_k = (B^T (I - W)^T P (I - W) B)^{-1} B^T (I - W)^T P (I - W) L. \tag{14}$$

2.4. Weighted Semiparameter Model. In model (7), parameter component is weighted α; nonparameter component is weighted β. Then the residual error of estimate value is

$$\widetilde{V_k} = \alpha(B_k X - L_k) + \beta \widehat{S}(t_k, X). \tag{15}$$

If B is nonsingular, the weighted semiparametric least squares estimation of X is

$$\widehat{X}_k = (B^T (\alpha I - \beta W)^T P(\alpha I - \beta W) B)^{-1} B^T (\alpha I - \beta W)^T \\ \times P(\alpha I - \beta W) L, \tag{16}$$

where $\alpha + \beta = 2$.

The weights α and β are adjusted appropriately based on the error of mean square. The formula of mean square error is shown as follows:

$$D = \left[\sum_{i=1}^{n} (X_{xi} - X_i)^2 \right]^{1/2}. \tag{17}$$

Step length is equal to 0.1. α go from 0 to 2 per step and β reduce from 2 to 0. There are one-to-one match from α, β to mean square error. The most appropriate weights correspond to the minimal mean square error.

When $\alpha = \beta = 1$, the weighted semiparameter model becomes semiparameter model. When $\alpha = 2, \beta = 0$, the weighted semiparameter model becomes LS model. In this paper step length is 0.1. The weights α and β are obtained using the MMSE principle.
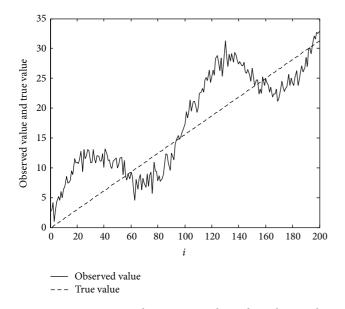
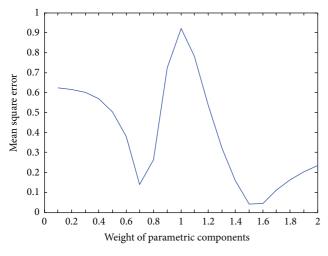FIGURE 1: Data comparison between true value and simulation value.

FIGURE 2: Mean square error with different weight.

3. Experiment Results and Analyses

3.1. Simulated Data. In this case, $L = BX + s + \Delta$, $B = (b_{i,j})_{100 \times 1}$, $n = 200$, $i = 1, 2, \ldots, n$, $b_i = t_i$, true value $X = 5$, $s = 6\sin(2t_i) + 2$, $t_i = (2(i-1)\pi)/n$, and $\Delta : N(0, 1)$. P is the weight matrix of observation values and $P = I$.

Figure 1 shows the data comparison between true value and simulation value.

In semiparameter and weighted semiparameter model, $K(\cdot) = \begin{cases} 1/2 & |x| \leq 1 \\ 0 & |x| > 1 \end{cases}$, where $K(\cdot)$ is kernel weighted function. Window width is 1.

Figure 2 shows the mean square error (MSE) with different weight.

So in this case $\alpha = 1.5$, $\beta = 0.5$.

Table 1 shows the estimate value and the gap between true value and simulation value.
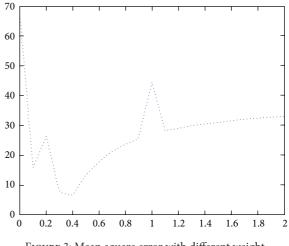

FIGURE 3: Mean square error with different weight.

From the value in Table 1, weighted semiparameter estimate model is the best. Using the weighted semiparameter estimate model, the estimate value is the nearest to the true value. The gap between true value and simulation value is significantly smaller than that in other models.

3.2. Data of Leveling Network. The data of leveling network comes from [15, page 109]. There are four points and five routes in leveling network. The condition number of coefficient matrix of equations is 8.4879e + 004. Its ill condition is not serious. So ridge estimate is not used. The weight kernel function is probability density function of standard normal distribution. Window width h equals 1.

In weighted semiparameter estimate model, mean square error is changing with weight α. Figure 3 shows the change process. It is evident from Figure 3 that $\alpha = 0.4$, $\beta = 1.6$ are the best choice.

Figure 4 shows the estimated residuals of different methods.

From Figure 4, the residual errors of weighted semiparameter estimate model and LS estimate are smaller than those of semiparameter estimate model. The model that has the least fluctuation is weighted semiparameter estimate.

Table 2 shows the mean square errors of three models.

The MSE follows weighted semiparameter—semiparameter—LS. The result of experiment indicates that the new method is more stable than others.

To summarize, according to the above experiments, weighted semiparameter estimate model is superior to other three estimate models.

4. Conclusions

A weighted semiparameter estimate model is proposed. The parameter components and nonparameter components are weighted by the characters of different data. Simulating data and leveling network data are used to test the model. Two indexes, namely, estimated residuals and MSE, are applied to analyze and verify the experimental result. The mean

TABLE 1: The estimate value and the gap between true and simulation value.

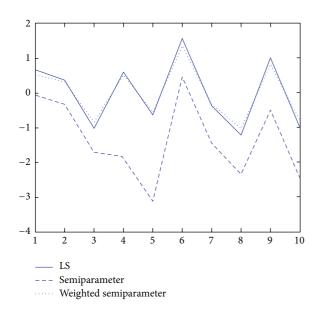|	True value	LS	Ridge estimate ($k = 0.7$)	Semiparameter	Weighted semiparameter $\alpha = 1.5, \beta = 0.5$
Estimate value	5	5.2350	5.2336	4.0787	4.9581
Gap	0	0.2350	0.2336	0.9213	0.0419

FIGURE 4: Estimated residuals of different methods.

TABLE 2: Mean square errors of three models.

	LS	Semiparameter	Weighted semiparameter
MSE	8.4477	30.4628	6.1035

square error and the range of residual error are the smallest. So weighted semiparameter estimate model has a better performance in effectiveness and high precision.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] X. Wang, *Theory and Application of Parameter Estimation of Nonlinear Model*, Wuhan University press, Wuhan, China, 2002.

[2] W. Xu, T. F. Coleman, and G. Liu, "A secant method for nonlinear least-squares minimization," *Computational Optimization and Applications*, vol. 51, no. 1, pp. 159–173, 2012.

[3] J. Xu, X. Qing, and G. Liu, *Control the Direction of the Theory of Least Square Method*, Surveying and Mapping Press, Beijing, China, 2010.

[4] A. Xiao, "Some algorithms of nonlinear least squares," *Mathematical Theory and Applications*, vol. 24, no. 2, pp. 86–90, 2004.

[5] J. M. Robins, A. Rotnitzkya, and L. P. Zhao, "Estimation of regression coefficients when some regressors are not always observed," *Journal of the American Statistical Association*, vol. 89, no. 427, pp. 846–866, 1994.

[6] J. Jan, S. van Bellegem, and V. Anne, "Iterative regularization in nonparametric instrumental regression," *Journal of Statistical Planning and Inference*, vol. 143, pp. 24–29, 2013.

[7] J. Zhang, Y. Peng, and O. Zhao, "A new semiparametric estimation method for accelerated hazard model," *Biometrics*, vol. 67, no. 4, pp. 1352–1360, 2011.

[8] L. Wang, L. D. Brown, and T. Cai, "A difference based approach to semiparametric partial linear model," Tech. Rep., 2007.

[9] G. Liu, Y. Jiang, and H. Tao, "Nonlinear least squares adjustment by parameters," *Acta Geodaetica et Cartographica Sinica*, vol. 27, no. 3, pp. 224–230, 1998.

[10] Z. Wang and H. Tao, "Determining the ridge parameter in a ridge estimation using l-curve method," *Geomatics and Information Science of Wuhan University*, vol. 29, no. 3, pp. 235–238, 2004.

[11] P. C. Hansen, "Analysis of discrete Ill-posed problems by means of l-curve," *SIAM Review*, vol. 34, no. 4, pp. 561–580, 1992.

[12] E. Haber and D. Oldenburg, "A GCV based method for nonlinear ill-posed problems," *Computational Geosciences*, vol. 4, no. 1, pp. 41–63, 2000.

[13] M. Marlene, *Semiparametric Extensions to Generalized Linear Models*, Schrift zur Habilitation im Fach Statistik, Berlin, Germany, 2000.

[14] J. Langford, *Learning Theory*, Springer, Berlin, Germany, 2005.

[15] B. Tao, *Error Theory and Survey Adjustment*, Wuhan University Press, Wuhan, China, 2007.