Two Stage Estimation in Semiparametric Model

Lihong Jin
Department of Basic, City College Wuhan University of Science and Technology, Wuhan, 430083, China

ABSTRACT: Considering the semi-parametric surveying adjustment models. Firstly, the estimators of parameters and nonparametric are derived by using the kernel function and least square method. Secondly, discuss the approach of bandwidth choice in the semi-parametric surveying adjustment models. Finally, a simulated adjustment problem is constructed to explain this method. The new method presented in this paper shows an effective way of solving the problem; the estimated values are nearer to their theoretical ones than the adjustment method of least squares.

KEYWORDS: Semiparametric adjustment models; Two-stage estimation; Kernel function; Model error

1 INTRODUCTION AND MOTIVATION

Stochastic factors and certain factors often coexist in survey data processing, conventional measurement adjustment adopts a stochastic model that just treats random error and does not treat certain systematic errors. However, systematic error surely exists in observation. If it is ignored, the adjusted result would be biased. When the model errors are minute as compared with random errors, omitting model errors does not influence the estimated value of the parameters seriously. When the model error is bigger, it will have a bad influence on parameter estimation, even resulting in false conclusions. It is generally assumed that genuine observations can be expressed by the functions depending on a group of parameters. Usually, the functional form of the regression model is assumed to be known. The problem is reduced to estimating a set of unknown parameters; in this case the observations are completely parameterized. Actually, it is difficult to describe the observations correctly by using several parameters.

In many applications, there is not always evidence of a linear relationship. Therefore, research for more flexible models is needed. In the 1980s, statistics brought forward semi-parametric regression models comprising a parametric part and a non-parametric part. It has some significant merits over parametric models in dealing with complicated relationships between observations and estimated variables and offering useful information. This model has been discussed by a number of authors. e.g. Fischer and Hegland (1999), Green and Silverman (1994), Shi and Tsai (1999), Sugiyama and Ogawa (2002), Aerts and Claeskens (2002), Ruppert (2002), Wang and Rao (2002), David (2003). This paper introduce the semi-parametric regression method for surveying adjustment, and to develop a suitable method that can estimate the unknown parameters and can extract model errors simultaneously.

2 THE MATHEMATICAL MODEL

The conventional survey adjustment is

\[ L_i = b_j x + \Delta_i \]  

If the model include systematic error, in order to separate systematic errors form model errors. Assume that responses \( L_i \) include non-parametric \( s(t_i) \), \( i = 1,2,\ldots,n \), then Eq.(1) can be rewritten as:

\[ L_i = b_j x + s(t_i) + \Delta_i \]  

Where \( b_j = (b_{i1}, b_{i2}, \ldots, b_{in}) \) is a vector of non-random design points, \( B = (b_1, b_2, \ldots, b_n)^T \) is of full rank with \( \text{rank}(B) = p \), that is, its columns are linearly independent, \( x = (x_1, x_2, \ldots, x_p)^T \) is a vector of unknown parameters, \( s = (s_1, s_2, \ldots, s_n)^T \) is used to describe vector of model errors. \( \Delta = (\Delta_1, \Delta_2, \ldots, \Delta_n)^T \) is an uncorrelated random
error vector with \( E(\Delta) = 0 \) and \( D_\Delta = \sigma^2 Q \), where \( \sigma^2 \) is the unit-weighted variance.

The error equation associated with Eq. (2) is
\[
V = B\hat{x} + \hat{s} - L
\]  
(3)

3 ESTIMATION METHOD AND MODELS
SOLUTION

Two stage estimate method attempts to estimate the parameter \( x \) and non-parametric \( s \) separately. More precisely, it can be derived from the following two-stage procedure.

Stage 1. For fixed \( x \), let \( \hat{\alpha} = \frac{1}{n} \sum_{i=1}^{n} s(t_i) \) and \( \hat{\varepsilon}_i = s(t_i) - \alpha + \Delta_i \), corresponding to Eq. (2), its equations is
\[
L_i = b_i x + \alpha + \hat{\varepsilon}_i, \quad i = 1, 2, \cdots, n
\]  
(4)

Where \( \hat{\varepsilon}_1, \hat{\varepsilon}_2, \cdots, \hat{\varepsilon}_n \) are zero mean, uncorrelated random variables having common infinite variance. To simplify the notation, we will let \( L = (L_1, L_2, \cdots, L_n)^T \), \( 1_n = (1, 1, \cdots, 1)^T \), \( \hat{x}^* \) and \( \hat{\alpha}^* \) be the least squares estimator of \( x \) and \( \alpha \). Corresponding to Eq. (4), its error equation is
\[
V_i = B\hat{x}^* + 1_n \hat{\alpha}^* - L
\]  
(5)

Based on the ordinary least squares methodology, construct the Lagrange function:
\[
\Phi_1 = V_i^T P V_i + 2\lambda_i^T (B\hat{x}^* + 1_n \hat{\alpha}^* - L - V_i)
\]  
(6)

Where \( P = Q^{-1} \) is symmetric positive-definite matrix, and also the weighted matrix of the observation \( L \) and \( \lambda_i \) is a Lagrange constant. Letting \( \frac{\partial \Phi_1}{\partial V_i} = 0 \), \( \frac{\partial \Phi_1}{\partial \hat{x}^*} = 0 \) and \( \frac{\partial \Phi_1}{\partial \hat{\alpha}^*} = 0 \),

Respectively, we obtain
\[
\lambda_i = P V_i
\]  
(7)

\[
B^T \lambda_i = 0
\]  
(8)

\[
\lambda_i^T \lambda_i = 0
\]  
(9)

By simple calculus, it follows that Eq. (6) is minimized when \( \hat{x}^* \) and \( \hat{\alpha}^* \) satisfy the block matrix equation:
\[
\begin{pmatrix}
B^T PB & B^T P1_n \\
1_n^T PB & 1_n^T P1_n
\end{pmatrix}
\begin{pmatrix}
\hat{x}^* \\
\hat{\alpha}^*
\end{pmatrix} =
\begin{pmatrix}
B^T PL \\
1_n^T PL
\end{pmatrix}
\]  
(10)

Eq. (10) forms a system of \( p + 1 \) equations: this is typically very large, and it may not be convenient, or even practical, to solve this system directly. Fortunately, this is not necessary. One approach is to re-write Eq. (10) as the pair of simultaneous matrix equations
\[
B^T PB\hat{x}^* + B^T P1_n \hat{\alpha}^* - B^T PL = 0
\]  
(11)

\[
1_n^T PB\hat{x}^* + 1_n^T P1_n \hat{\alpha}^* - 1_n^T PL = 0
\]  
(12)

From Eqs. (11), (12), we obtain
\[
\hat{x}^* = (B^T PB)^{-1} B^T P(L - 1_n \hat{\alpha}^*)
\]  
(13)

\[
\hat{\alpha}^* = (1_n^T P1_n)^{-1} 1_n^T P(L - B\hat{x}^*)
\]  
(14)

It turns out that we can use an alternation between Eqs. (13) and equation (14), solving repeatedly for \( \hat{x}^* \) and \( \hat{\alpha}^* \) respectively. This procedure is sometimes known as back fitting (Green and Silverman 1994). In practice, the convergence can be very slow. An alternative approach is a direct method. Let \( d_n = 1_n^T (P - PB(B^T PB)^{-1} B^T P)1_n \), substituting Eq. (13) into Eq. (14), we get the first estimate \( \hat{\alpha}^* \) of \( \alpha \)
\[
\hat{\alpha}^* = d_n^{-1} 1_n^T (P - PB(B^T PB)^{-1} B^T P)L
\]  
(15)

Substituting Eq. (15) into Eq. (13), we get the first estimator \( \hat{x}^* \) of \( x \)
\[
\hat{x}^* = (B^T PB)^{-1} B^T P(L - 1_n \hat{\alpha}^*)
\]  
(16)

Stage 2. Substituting \( \hat{x}^* \) into model Eq. (2), the model Eq. (2) can be considered as a non-parametric regression model
\[
L_i - b_i \hat{x}^* = s(t_i) + \Delta_i, \quad i = 1, 2, \cdots, n
\]  
(17)

Therefore, \( s(t_i) \) can be estimated by the non-parametric kernel method
\[
\hat{s}(t_i) = \sum_{j=1}^{n} K_h(t_i, t_j) (L_j - b_j \hat{x}^*)
\]  
(18)

Where the weight \( K_h(\cdot) \) is associated with a kernel function \( K(\cdot) \) and its band width \( h = h_n > 0 \). For asymmetric kernel function \( K(\cdot) \), the weight \( K_h(\cdot) \) is taken to be
\[
K_h(t, t') = \frac{1}{h} K \left( \frac{t - t'}{h} \right)
\]  
(19)

Substituting \( \hat{s}(t_i) \) for \( s(t_i) \) in model Eq. (2) gives
\[
L_i = b_i x + \hat{s}(t_i) + \Delta_i, \quad i = 1, 2, \cdots, n
\]  
(20)

Its error equation is
\[ V_2 = B \hat{x} + \hat{s} - L \]  

(21)

Based on the ordinary least squares methodology, we obtain

\[ \Phi_2 = V_2^T PV_2 + 2 \lambda_2 T (B \hat{x} + \hat{s} - L - V_2) \]  

(22)

Letting \( \frac{\partial \Phi_2}{\partial V_2} = 0 \), and \( \frac{\partial \Phi_2}{\partial x} = 0 \), respectively. We obtain

\[ \lambda_2 = PV_2 \]  

(23)

\[ B^T \lambda_2 = 0 \]  

(24)

Substituting Eq. (23) into Eq. (24), and considering (21)

\[ B^T PBx + B^T P\hat{s} - B^T PL = 0 \]  

(25)

Since \( B^T PB \) is a positive matrix, we get the final estimate of \( x \)

\[ \hat{x} = (B^T PB)^{-1} B^T P(L - \hat{s}) \]  

(26)

Considering Eq. (18), we also obtain the final estimate of \( s \)

\[ \hat{s}(t) = \sum_{i=1}^{n} K_h(t, t_i)(L_i - b_i \hat{x}) \]  

(27)

The ordinary least-squares fit to the data is

\[ \hat{L} = (W + (I - W)B(B^T PB)B^T \hat{P})L \equiv J(h)L \]  

(28)

4 BANDWIDTH SELECTION

The residual sum of squares (RSS) of a model is a measure of predictive ability, since a residual is the difference between an observation of a response and its fitted or predicted value, \( e_i \equiv L_i - \hat{L}_i \). There is a simple remedy to this problem: when predicting \( L_i \), use all the observations except the \( i \)th one. Thus, define \( \hat{L}^{(i)}(t_i, h) \) is the semi-parametric regression estimator applied to the data but with \( (t_i, L_i) \) deleted. Then, let \( e^{(i)}_i \equiv L_i - \hat{L}^{(i)}(t_i, h) \) be the \( i \)th deleted residual. This method is the technique of model validation that splits the data into two disjoint sets, fits the model to one set, predicts the data in the second set using only the fit to the first set, and then compares these predictions to the actual observations in the second set. This “leaving one out” strategy is a way of guarding against the wiggly answer that RSS gives. The choice of \( h \) is the one that minimizes PRESS over \( h > 0 \).

The PRESS criterion is not as difficult to calculate as it might first appear. One does not need to fit the model \( n \) times, thanks to an important identity. Let \( J_{ii}(h) \) be the \( i \)th diagonal element of the hat matrix \( J(h) \). Then the \( i \)th deleted residual is related to the \( i \)th ordinary residual by

\[ L_i - \hat{L}^{(i)}(t_i, h) = \frac{L_i - \hat{L}_i}{1 - J_{ii}(h)} \]  

(29)

Using Eq. (29), we obtain

\[ PRESS(h) = \sum_{i=1}^{n} \left( \frac{L_i - \hat{L}_i}{1 - J_{ii}(h)} \right)^2 \]  

(30)

Thus, the PRESS can be computed using only ordinary residuals and the diagonal elements of the hat matrix.

5 SIMULATION EXPERIMENT

We compared the semi-parametric approach with the ordinary least squares adjustment approach. For a simulated simple semi-parametric model \( L = BX + S + \Delta \), where \( X = [2,3]^T \), \( B = (b_i)_{100 \times 2} \), \( b_{1i} = -2 + i/20 \), \( b_{2i} = -2 + (i/20)^2 \), model errors \( s_i = 9 \sin(t_i \pi) \), \( t_i = i/100 \), \( i = 1,2,\ldots,200 \). The observation error vector \( \Delta \) is composed of 200 stochastic data obeying the standard normal distribution, which are \( N(0,1) \) distributed. In Figure 1, the smooth curve shows the true values \( BX \); the sawtooth curve is the observation \( L \).

Choose the weighted matrix \( P \) is an identity matrix. Using the parametric model and the least squares adjustment approach, we obtain

\[ \hat{X} = (B^T PB)^{-1} B^T PL \equiv (-3.1194, 3.3611)^T \]

An incorrect result may be given with the conventional technique when the model error exits. We use the quartic kernel \( K(t) = 15/16((1 - t^2)^3) \),

Figure 1. The Observations and Their True Values.
the results for different bandwidth are presented in Table 1. Table 1 lists the bandwidth $h$ and estimation of parametric component $\hat{X}$.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$X$</th>
<th>$h$</th>
<th>$X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.11</td>
<td>(1.1938, 2.6599)</td>
<td>0.20</td>
<td>(1.9264, 2.8025)</td>
</tr>
<tr>
<td>0.12</td>
<td>(1.3695, 2.6545)</td>
<td>0.21</td>
<td>(1.9021, 2.8390)</td>
</tr>
<tr>
<td>0.13</td>
<td>(1.5196, 2.6553)</td>
<td>0.22</td>
<td>(1.8598, 2.8783)</td>
</tr>
<tr>
<td>0.14</td>
<td>(1.6451, 2.6620)</td>
<td>0.23</td>
<td>(1.8003, 2.9200)</td>
</tr>
<tr>
<td>0.15</td>
<td>(1.7468, 2.6742)</td>
<td>0.24</td>
<td>(1.7245, 2.9638)</td>
</tr>
<tr>
<td>0.16</td>
<td>(1.8254, 2.6914)</td>
<td>0.25</td>
<td>(1.6332, 3.0092)</td>
</tr>
<tr>
<td>0.17</td>
<td>(1.8819, 2.7132)</td>
<td>0.26</td>
<td>(1.5274, 3.0560)</td>
</tr>
<tr>
<td>0.18</td>
<td>(1.9171, 2.7392)</td>
<td>0.27</td>
<td>(1.4079, 3.1038)</td>
</tr>
<tr>
<td>0.19</td>
<td>(1.9316, 2.7690)</td>
<td>0.28</td>
<td>(1.2755, 3.1523)</td>
</tr>
</tbody>
</table>

By calculation Eq.(30), the selected bandwidth is $h=0.2096$ and estimation of parametric component is

$$\hat{X} = (1.9034, 2.8375)^T$$

The estimation $\hat{X}$ approximate to the true values of $X$, the model errors is very close to the value of $S$, thus it can be seen that our method is successful.

### 6 CONCLUSION

In many applications the parametric model itself is at best an approximation of the true one, and the search for an adequate model from the parametric family is not easy. When there are no persuasive models available, the family of semi-parametric regression models can be considered as a promising extension of the parametric family. Semi-parametric regression models reduce complex data sets to summaries that we can understand. Properly applied, they retain essential features of the data while discarding unimportant details, and hence they aid sound decision-making.

From the above, it is difficult for the adjustment method of least squares to detect systematic errors in the data processing model, but semi-parametric regression model is valid. Added only a few additional parameters the model cannot express the complicated model errors, which will bring bad influence to the estimation of the parameters without considering the model errors shown in the examples. Introduce the semi-parametric regression theory for surveying adjustment, may be reduce the error of semi-parametric estimators by choosing an appropriate value for the kernel function and bandwidth.

On the other hand, there exit a lot of problems to discuss, for instance: how to choose kernel function and bandwidth? It is one of successful keys and needs to be further explored, and is under investigation.

### REFERENCES