Research Article

Two-Stage Method Based on Local Polynomial Fitting for a Linear Heteroscedastic Regression Model and Its Application in Economics

Liyun Su, Yanyong Zhao, and Tianshun Yan

School of Mathematics and Statistics, Chongqing University of Technology, Chongqing 400054, China

Correspondence should be addressed to Liyun Su, cloudhopping@163.com

We introduce the extension of local polynomial fitting to the linear heteroscedastic regression model. Firstly, the local polynomial fitting is applied to estimate heteroscedastic function, then the coefficients of regression model are obtained by using generalized least squares method. One noteworthy feature of our approach is that we avoid the testing for heteroscedasticity by improving the traditional two-stage method. Due to nonparametric technique of local polynomial estimation, we do not need to know the heteroscedastic function. Therefore, we can improve the estimation precision, when the heteroscedastic function is unknown. Furthermore, we focus on comparison of parameters and reach an optimal fitting. Besides, we verify the asymptotic normality of parameters based on numerical simulations. Finally, this approach is applied to a case of economics, and it indicates that our method is surely effective in finite-sample situations.

1. Introduction

The heteroscedasticity in classical linear regression model is defined by the variances of random items which are not the same for different explanatory variables and observations. Heteroscedasticity often occurs in data sets in which there is a wide disparity between the largest and smallest observed values. The larger the disparity between the size of observations in a sample, the larger the likelihood that the error term observations associated with them will have different variances and therefore be heteroscedastic. That is, we would expect that the error term distribution for very large observations might have a large variance, but the error term distribution for small observations might have a small variance. Besides, researchers have observed that heteroscedasticity is usually found in cross-sectional data rather than in

time series data [1, 2]. In cross-sectional data we generally deal with members of a population at a given point in time, such as individual consumers or their families; firms; industries; or geographical subdivisions, such as small, medium, or large firms, or low, medium, or high income. Besides, this study is also more common in economics, such as the relationship of sale and research and development, and profit in a certain year.

When there is a heteroscedasticity in a linear regression model [3–6], we can apply parametric methods, also the heteroscedastic approaches can be applied, such as Park test, White test and so on. Another heteroscedastic method is shown in [7]. When parametric approaches are applied, estimations of parameters we obtained by ordinary least squares estimation (OLS) are still linear and unbiased. However, the efficiency is bad [8–10]. This could lead to a uncorrect statistical diagnosis for the parameters' significance test. Similarly, it unecessarily enlarges the confidence interval when we estimate the parameter interval. Besides, the accuracy of predictive value may lower when we estimate with the regression model that we obtained. In order to solve the problem above, we can use generalized least squares estimation (GLS) when the covariance matrix of the random items is known. If it is unknown, we usually use two-stage least squares estimate, that is, we first estimate variances of the residual error, and then the generalized least squares estimator is used to obtain the coefficients of the model by using the estimate of variances of the random items [11, 12]. However, the traditional estimation method is that we suppose the residual error variances as a certain parametric model. In this paper, we try to applying local polynomial fitting to random item variances as the first step, and then GLS is used to estimate the coefficients of the model. A problem cannot be neglected is that the existence of inverse of regressor matrix transpose multiplied by regressor matrix is necessary for the existence of the unique estimator when applying the least squares method or their generalized versions. In the generalized versions, because kernel function is a symmetric probability density function with bounded support, the weighted matrix based on kernel function is positive definite, generally being also symmetric as done in usual theoretical and practical selections. In other words, the regressor matrix has to be full column rank and weighted matrix has to be nonsingular. The issues have been discussed in many papers used in theoretical issues and many applications, see [13–17]. On the one hand, because of local polynomial fitting's various nice statistical properties, the estimations obtained with this technology also possess the same good statistical properties [12, 18, 19]. On the other hand, we exploit a heteroscedastic regression model rather than the artificial structure of heteroscedasticity. Then, we can directly get the heteroscedastic function based on the nonparametric technique, which shows the relationship between variance function of random items and explanatory variables from regression results. Thus, it is unnecessary to test heteroscedasticity of the model. Particularly, the estimated value by local polynomial fitting is more accurate than that by the traditional method. Besides, we study variance function fitting when parameters change and reach the optimal parameters. Finally, a case of economics is cited in order to show that our method is indeed effective in finite-sample situations.

The rest of this paper is organized as follows. In Section 2 we construct local polynomial regression: in Section 2.1 we talk about local polynomial fitting, in Section 2.2 we study estimation with Parameters selections. Section 3 contains two-stage method with local polynomial fitting. In Section 4, we do the simulations on a given model and study the fittings under different parameters. In Section 5, we collect some real data and use the local polynomial estimating the coefficients. We draw the results in Section 6.

2. Local Polynomial Regression

Local polynomial regression [20–23] is a widely used nonparametric technique. Local polynomial fitting is an attractive method both from theoretical and practical point of view. Local polynomial method has a small mean squared error compared with the Nadaraya-Watson estimator which leads to an undesirable form of the bias and the Gasser-Muller estimator which has to pay a price in variance when dealing with a random design model. Local polynomial fitting also has other advantages. The method adapts to various types of designs such as random and fixed designs, highly clustered and nearly uniform designs. Furthermore, there is an absence of boundary effects: the bias at the boundary stays automatically of the same order as the interior, without use of specific boundary kernels. The local polynomial approximation approach is appealing on general scientific grounds: the least squares principle to be applied opens the way to a wealth of statistical knowledge and thus easy generalizations. In this section, we briefly outline the idea of the extension of local polynomial fitting to linear regression.

2.1. Local Polynomial Fitting

Consider the bivariate data $(X_1, Y_1), \ldots, (X_n, Y_n)$, which form an independent and identically distributed sample from a population (X, Y). Of interest is to estimate the regression function $m(x_0) = E(Y \mid X = x_0)$ and its derivatives $m'(x_0), m''(x_0), \ldots, m^{(p)}(x_0)$. To help us understand the estimation methodology, we can regard the data as being generated from the model

$$Y = m(X) + \sigma(X)\varepsilon, \tag{2.1}$$

where $E(\varepsilon) = 0$, $\mathrm{Var}(\varepsilon) = 1$, and X and ε are independent. However, this location-scale model assumption is not necessary for our development, but is adopted to provide intuitive understanding. We always denote the conditional variance of Y given $X = x_0$ by $\sigma_2(x_0)$ and the marginal density of X, that is, the design density, by $f(\cdot)$.

Suppose that the $(p+1)$th derivative of $m(x)$ at the point x_0 exists. We then approximate the unknown regression function $m(x)$ locally by a polynomial of order p. A Taylor expansion gives, for x in a neighborhood of x_0,

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0) + \frac{m''(x_0)}{2!} + \cdots + \frac{m^{(p)}}{p!}(x_0)(x - x_0)^p. \tag{2.2}$$

This polynomial is fitted locally by a weighted least squares regression problem: minimize

$$\sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j \right\}^2 K_h(X_i - x_0), \tag{2.3}$$

where h is a bandwidth controlling the size of the local neighborhood, and $K_h(\cdot) = K(\cdot/h)/h$ with K a kernel function assigning weights to each datum point. Throughout this paper we assume that K is a symmetric probability density function with bounded support, although this technical assumption can be relaxed significantly [24–26].

Denote by $\hat{\beta}_j, j = 0, \ldots, p$, the solution to the least squares problem (2.3). It is clear from the Taylor expansion in (2.2) that $\hat{m}_\nu(x_0) = \nu!\hat{\beta}_\nu$ is an estimator for $m^{(\nu)}(x_0), \nu = 0, 1, \ldots, p$. To estimate the entire function $m^{(\nu)}(\cdot)$ we solve the above weighted least squares problem for all points x_0 in the domain of interest.

It is more convenient to work with matrix notation. Denote by $\mathbf{X}$ the design matrix of problem (2.3):

$$\mathbf{X} = \begin{bmatrix} 1 & (X_1 - x_0) & \cdots & (X_1 - x_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & (X_n - x_0) & \cdots & (X_n - x_0)^p \end{bmatrix}, \tag{2.4}$$

and let

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \qquad \hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}. \tag{2.5}$$

Further, let $\mathbf{W}$ be the $n \times n$ diagonal matrix of weights:

$$\mathbf{W} = \mathrm{diag}\{K_h(X_i - x_0)\}. \tag{2.6}$$

Then the weighted least squares problem (2.3) can be written as

$$\min_\beta (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{W}(\mathbf{y} - \mathbf{X}\beta), \tag{2.7}$$

with $\beta = (\beta_0, \ldots, \beta_p)^T$. The solution vector is provided by weighted least squares theory and is given by

$$\hat{\beta} = \left(\mathbf{X}^T \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}\mathbf{y}, \tag{2.8}$$

where the regressor matrix $\mathbf{X}$ has to be full column rank and the weighted matrix $\mathbf{W}$ has to be nonsingular. If such an inverse in (2.8) does not exist, a method of ridge regression can be adopted to solve this problem [27]. Furthermore, we can get the estimation $\hat{m}(x_0)$,

$$\hat{m}(x_0) = E_1 \left(\mathbf{X}^T \mathbf{W} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}\mathbf{y}, \tag{2.9}$$

where E_1 is a column vector (the same size of β) with the first element equal to 1, and the rest equal to zero, that is, $E_1 = (1, 0, \ldots, 0)_{1 \times (p+1)}$.

Computing the $\hat{\beta}$ will suffer from large computational cost. We can use the recursive least squared method to reduce the computation complexity, and it is very powerful

especially in the local polynomial fitting problems. There are several important issues about the bandwidth, the order of local polynomial function and the kernel function which have to be discussed. The three problems will be presented in Section 2.2.

2.2. Parameters Selections

To implement the local polynomial estimator, one need to choose the order p, the kernel K and the bandwidth h. These parameters are of course related to each other.

First of all, the choice of the bandwidth parameter h is considered, which plays a rather crucial role. A too large bandwidth under-parametrizes the regression function [28, 29], causing a large modeling bias, while a too small bandwidth over-parametrizes the unknown function and results in noisy estimates. The basic idea is to find a bandwidth h that minimizes the estimated mean integrated square error (MISE):

$$h_{\text{opt}} = \arg\min_h \int \left\{ (\text{Bias}(\widehat{m}(x) \mid X))^2 + \text{Var}(\widehat{m}(x) \mid X) \right\} dx, \tag{2.10}$$

where $X = (X_1, \ldots, X_p)$, the asymptotic bias and the asymptotic variance are denoted in Lemma A.1 of the appendix. Then we can find an asymptotically optimal constant bandwidth given by

$$h_{\text{opt}} = c_{\nu,p}(K) \left[\frac{\int \sigma^2(x)/f(x)}{\int \{m^{(p+1)}(x)dx\}} \right]^{1/(2p+3)} n^{-1/(2p+3)}, \tag{2.11}$$

where, $c_{\nu,p}(K)$ is a constant which relates to the kernel function and the order of the local polynomial, $f(x)$, is the density function of $m(\cdot)$. However, this ideal bandwidth is not directly usable since it depends on unknown functions.

Another issue in local polynomial fitting is the choice of the order of the polynomial. Since the modeling bias is primarily controlled by the bandwidth, the issue is less crucial however. For a given bandwidth, h, a large value of p would expectedly reduce the modeling bias, but would cause a large variance and a considerable computational cost. It is shown in [30] that there is a general pattern of increasing variability: for estimate $m^{(\nu)}(x_0)$, there is no increase in variability when passing from an even (i.e., $p - \nu$ even) $p = \nu + 2q$ order fit to an odd $p = \nu + 2q + 1$ order fit, but when passing from an odd $p = \nu + 2q + 1$ order fit to the consecutive even $p = \nu + 2q + 2$ order fit, there is a price to be paid in terms of increased variability. Therefore, even order fits $p = \nu + 2q$ are not recommended. Since the bandwidth is used to control the modeling complexity, we recommend the use of lowest odd order, that is, $p = \nu + 1$, or occasionally $p = \nu + 3$.

Another question concerns the choice of the kernel function K. Since the estimation is based on the local regression (2.3), no negative weight K should be used. As shown in [30], the optimal weight function is $K(z) = (3/4)(1 - z^2)_+$, the Epanechnikov kernel, which minimizes the asymptotic mean square error (MSE) of the resulting local polynomial estimators.

3. Two-Stage Method with Local Polynomial Fitting

Let the dependent variable x and the explanatory variable y fulfill the following regression model:

$$y_i = \beta_0 + \beta_1 x_i + \sigma(x)\varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{3.1}$$

where y_i are the observations and $x_i, \ i = 1, \ldots, n$ are independent variables. Denote

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix},$$

$$\mathbf{X}_x = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \tag{3.2}$$

Therefore, (3.1) can be abbreviated as

$$y = \mathbf{X}_x \beta + \sigma(x)\varepsilon. \tag{3.3}$$

Suppose that

(1) $E(\sigma(x)\varepsilon) = (E(\sigma(x)\varepsilon_1), \ldots, E(\sigma(x)\varepsilon_n))^T = (0, \ldots 0,)^T.$

(2) $\Sigma = E((\sigma(x)\varepsilon)(\sigma(x)\varepsilon)^T) = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, where $\sigma_i^2 = \operatorname{var}((\sigma(x)\varepsilon_i)), i = 1, 2, \ldots, n.$

(3) $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ are not all equal, that is, there is a heteroscedasticity in model (3.4). Therefore, GLS for β is

$$\widehat{\beta} = \left(\mathbf{X}_x^T \Sigma^{-1} \mathbf{X}_x \right)^{-1} \mathbf{X}_x^T \Sigma^{-1} y. \tag{3.4}$$

If covariance matrix Σ is known, the coefficients β can be estimated. Equation (3.4) is considered as the weighted least squares estimation (WLS) for β and it possesses nice qualities. However, how to estimate the Σ is still a problem. Therefore, the so-called two-stage method of estimation is used to solve the heteroscedasticity problem. The two-stage method based on local polynomial fitting can be depicted as follows: firstly, apply local polynomial

fitting to get the estimate $\widehat{\Sigma}$ for Σ, that is, $\widehat{\sigma}_i^2$ for σ_i^2, and then we can obtain the estimate $\widehat{\beta}$ for β by using (3.5). The estimator follows that

$$\widehat{\beta} = \left(\mathbf{X}_x^T \widehat{\Sigma}^{-1} \mathbf{X}_x\right)^{-1} \mathbf{x}_x^T \widehat{\Sigma}^{-1} y = \left(\sum_{i=1}^{n} \frac{1}{\widehat{\sigma}_i^2} x_i x_i^T\right)^{-1} \left(\sum_{i=1}^{n} \frac{1}{\widehat{\sigma}_i^2} x_i y_i\right). \tag{3.5}$$

Because of $E(\sigma(x_i)\varepsilon_i \mid x_i)^2 = \sigma_i^2$, we construct the following regression model in order to estimate σ_i^2,

$$(\sigma(x_i)\varepsilon_i)^2 = \sigma_i^2 + u_i, \quad Eu_i = 0, \ i = 1, 2, \ldots, n, \tag{3.6}$$

where u_i is the difference between $(\sigma(x)\varepsilon_i)^2$ and its expectation. Suppose that $b = (\mathbf{X}_x^T \mathbf{X}_x)^{-1} \mathbf{X}_x^T y$ is the OLS of model (3.4). Although the ordinary least squares estimate b is ineffective, it is still consistent. Therefore, the corresponding residuals hold that

$$e_i^2 = \left(x_i^T \beta + (\sigma(x)\varepsilon_i) - x_i^T b\right)^2$$
$$= (\sigma(x)\varepsilon_i)^2 + \left[x_i^T (\beta - b)\right]^2 + 2(\sigma(x)\varepsilon_i) x_i^T (\beta - b) \approx (\sigma(x)\varepsilon_i)^2. \tag{3.7}$$

Consequently, we can approximately get

$$e_i^2 = \sigma_i^2 + v_i, \quad E(v_i) = 0, \ i = 1, 2, \ldots, n. \tag{3.8}$$

It can be taken as a regression model, in which the variance function is regression function and the squared residuals e_i^2 are dependent variables. In order to estimate this model, parameter estimation method would usually be taken in some articles. In other words, they suppose $\sigma_i^2 = f(c^T, x_i)$, where the form of f is known and $c = (c_0, c_1)^T$ are the parameters to be estimated. Note that what we usually discuss about are $\sigma^2 = \sigma^2(c^T, x_i)$, $\sigma_i^2 = \sigma_i^2 \ln(c^T, x_i)$ and so on [31]. However, the discussion for these models requires the analysts to have a better understanding of the background in practical problems. As an example, variance of asset return is always in direct proportion with its past return. Since the variance function must be nonnegative, a nonparametric method is proposed to fit σ_i^2. This method can be depicted as follows. Then, a p-order local polynomial estimation for the variances function $\sigma^2(x)$ is obtained according to formula (2.2). Using the least squares method for the data around the local window, we can estimate the local intercept via minimizing

$$\sum_{i=1}^{n} \left[e_i^2 - \alpha_0 - \alpha_1(x - x_i) - \cdots - \alpha_p(x - x_i)^p\right]^2 K\left(\frac{x - x_i}{h}\right), \tag{3.9}$$

with respect to $\{\alpha_i\}_{i=0}^{p}$. Therefore, the solution vector can be written as

$$\alpha = \left(X_{xx}^T W_x X_{xx}\right) X_{xx}^T W_x e^2, \tag{3.10}$$

where the design matrix

$$
\mathbf{X}_{xx} = \begin{bmatrix}
1 & (x - x_1) & (x - x_1)^2 & \cdots & (x - x_1)^p \\
1 & (x - x_2) & (x - x_2)^2 & \cdots & (x - x_2)^p \\
\vdots & \vdots & \vdots & \cdots & \vdots \\
1 & (x - x_n) & (x - x_n)^2 & \cdots & (x - x_n)^p
\end{bmatrix}, \tag{3.11}
$$

the weighted matrix

$$
\mathbf{W}_x = \frac{1}{h} \begin{bmatrix}
K\left(\dfrac{x - x_1}{h}\right) \\
& K\left(\dfrac{x - x_2}{h}\right) \\
& & \ddots \\
& & & K\left(\dfrac{x - x_n}{h}\right)
\end{bmatrix}, \tag{3.12}
$$

and $e^2 = (e_1^2, e_2^2, \ldots, e_n^2)^T$. Consequently, the estimated variance function is $\hat{\sigma}^2(x) = E_1\alpha$. Finally, we can get two-stage estimate $\hat{\beta}$ for β by substituting estimate $\hat{\sigma}_i^2$ for σ_i^2 into (3.5), which has some wonderful statistics qualities, seeing Lemma A.2, Lemma A.3, and Theorem A.4 in the appendix.

4. Simulation and Analysis

In order to discuss the qualities of $\hat{\beta}$ under the limited sample, this section gives the following model, in which we do the comparison and study the fittings under different parameters. Considering the practical background which is applied to economics, we suppose the variance function of the following form. Besides, we study the fitting effects of variance in different parameters to obtain the best one.

Denote the linear model by

$$
y_i = 3.1 + 1.5x_i + \sigma(x_i)\varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{4.1}
$$

where x_i $(i = 1, \ldots, n)$ are independent variables, y_i $(i = 1, \ldots, n)$ are observations. Also, $E(\sigma(x_i)\varepsilon_i)(\sigma(x_i)\varepsilon_i)^T = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, and $\sigma_1^2, \ldots, \sigma_n^2$ are not all equal. Suppose that the variance function of the error term is $\sigma^2(x) = e^{0.6x}$.

Step 1. Firstly, obtain the estimation of two coefficients in model (4.1) with the ordinary least squares estimation. Secondly, calculate the squares of the residuals $e^2 = (e_1^2, \ldots, e_n^2)^T$. Thirdly, do the local polynomial regression based on the model (3.8). In this section, it is necessary to discuss how to choose parameters. Suppose that the range of x is $[-1, 5]$ and the kernel function is the Epanechnikov kernel $K(z) = (3/4)(1 - z^2)_+$. In addition, the criteria of selecting

Table 1: Values of MISE under different orders and bandwidths.

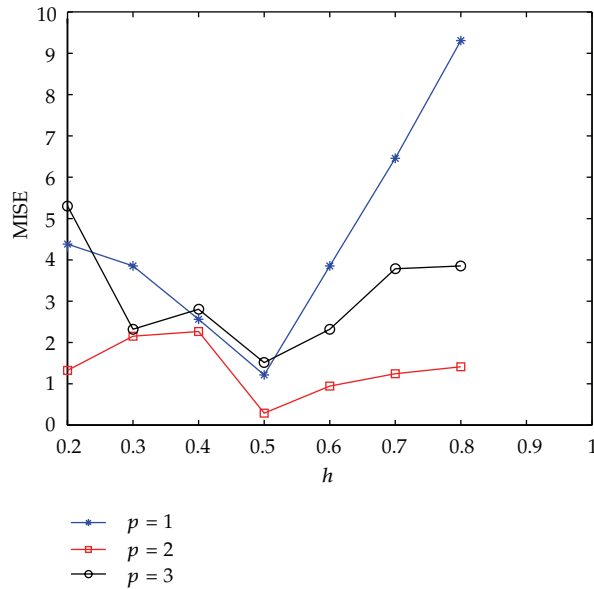| MISE	$h = 0.2$	$h = 0.3$	$h = 0.4$	$h = 0.5$	$h = 0.6$	$h = 0.7$	$h = 0.8$
$p = 1$	4.383078	3.852483	2.5617354	1.215800	3.853911	6.459762	9.307465
$p = 2$	1.325719	8.152896	2.265193	0.289220	0.945407	1.245113	1.413955
$p = 3$	5.303323	2.320999	2.805549	1.512751	2.322089	3.784113	3.854334

Figure 1: The scatter plot under different orders.

a bandwidth h is minimizing the mean integrated square error (MISE), which can be given by

$$\text{MISE} \equiv \int E\left\{\sigma^2(x) - \hat{\sigma}^2(x)\right\}^2 dx. \tag{4.2}$$

We could get values of MISE under different orders and bandwidths by 10000 replicates calculation with the above equation, where $n = 300$, see Table 1.

Furthermore, the scatter plot about h and MISE under different orders can be drawn, see Figure 1.

From Figure 1, it can be seen that values of MISE are the minimum when $h = 0.5$ whether $p = 1$, $p = 2$, and $p = 3$. Therefore, $h = 0.5$ is the optimal bandwidth. Further, value of MISE when $p = 2$ is the minimum among the above three, that is, $p = 2$ is the optimal order. It can be drawn the conclusion that the optimal parameters are $h = 0.5$ and $p = 2$, respectively. Figure 2 shows the fitting plot and residual plot for variance function after 10000 replicates, where $h = 0.5$, $p = 2$ and $n = 300$.

Step 2. Now we substitute $\hat{\sigma}_i^2$ which is obtained from Step 1 into model (3.5), and then we will get the estimation $\hat{\beta}$ of GLS, that is, $\hat{\beta}_0 = 3.1004$, $\hat{\beta}_1 = 1.4987$. Figure 3 depicts the histograms and asymptotic distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$, by which we do 10000 replicates with GLS and
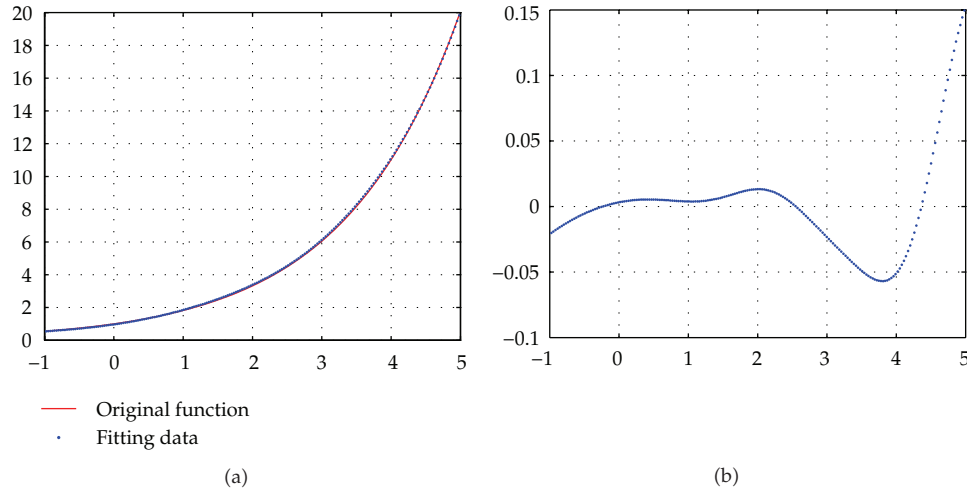
Original function
· Fitting data

(a)   (b)

Figure 2: The fitting plot and its residual plot when $h = 0.5$ and $p = 2$: (a) fitting plot, (b) residual plot.
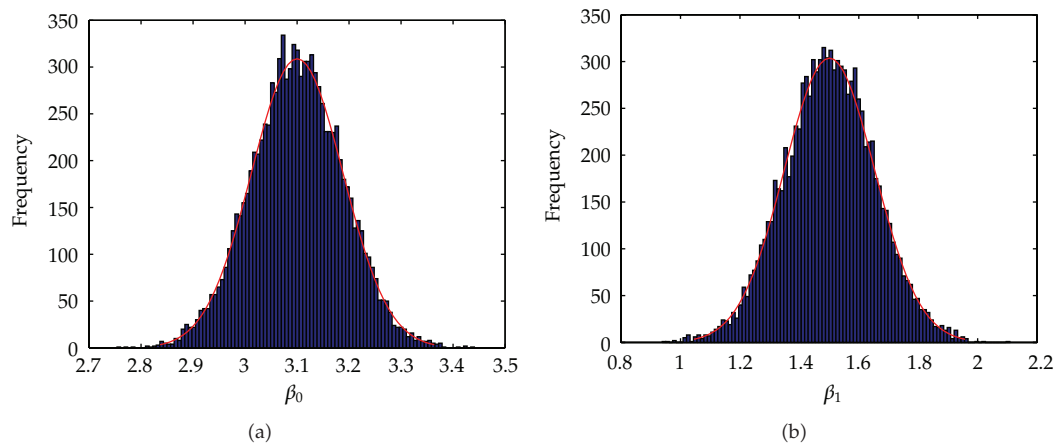

(a)   (b)

Figure 3: The histograms and asymptotic distributions of $\widehat{\beta}_0$ and $\widehat{\beta}_1$: (a) $\widehat{\beta}_0$, (b) $\widehat{\beta}_1$.

choose $n = 300$. It is easy to see from Figure 3 that the estimated distributions of parameters are subject to normal asymptotically. Besides, the OLS for β_0 and β_1 can easily be obtained, say, $\widehat{b}_0 = 3.1813$, $\widehat{b}_1 = 1.4394$. The fitted and true values for GLS and ones of OLS are listed in Table 1. Here, the relative error is defined by Re $= (|\widehat{\xi} - \xi|/\xi) \times 100\%$, where ξ and $\widehat{\xi}$ present true and fitted values, respectively. From the comparison in Table 2, we can conclude that the estimators for parameters by GLS are much better than those by OLS. Furthermore, curves of original and regression functions estimated by GLS and OLS are plotted together in order to demonstrate accuracy by GLS, see Figure 4. It is not difficult to see that the GLS regression curve is almost coincident with original curve, while there is a parent bias between OLS regression and original curve. Consequently, estimators for parameters by GLS are what we require.

Table 2: Two-stage estimates for β_0 and β_1 of GLS and OLS.

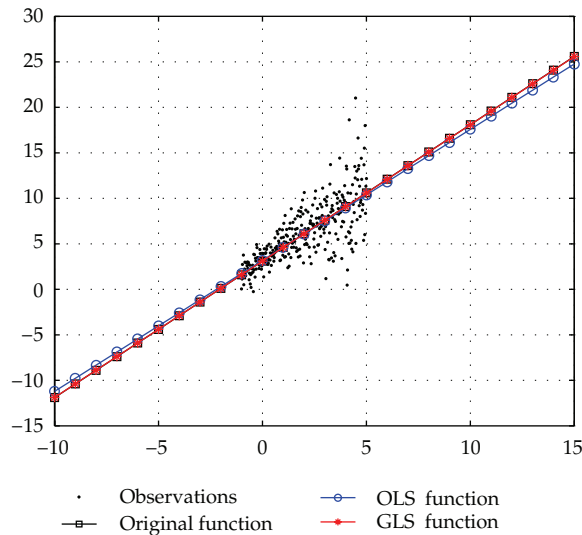| Parameters		β_0	β_1
True value		3.1	1.5
GLS	Fitted value	3.1004	1.4987
	Relative error	0.0129%	0.0867%
OLS	Fitted value	3.1813	1.4394
	Relative error	2.6226%	4.0400%

Figure 4: Curves of original and regression functions estimated by GLS and OLS.

5. Application

As an example of pure cross-sectional data with potential for heteroscedasticity, consider the data given in Table 3, which gives data on per capita consumption (Y) and per capita gross domestic product (GDP)(X) for 31 province or city in the People's Republic of China in 2008. Since the cross-sectional data presented in this table are quite heterogenous in a regression of per capita consumption (Y) on per capita GDP (X), heteroscedasticity is likely.

If we want to understand relationship between per capita consumption and per capita GDP, then the regression function is as follows:

$$Y = B_0 + B_1 X + u, \tag{5.1}$$

where B_0 and B_1 are regression coefficients and μ is a vector of random errors. The ordinary least squares (OLS) estimations for B_0 and B_1 can easily be obtained, that is, $\widehat{B}_0 = 679.72$ and $\widehat{B}_1 = 0.31117$. As follow, according to the known sample data, the regression plot with OLS can easily be drawn, seeing Figure 5. Not surprising, there is a positive relationship between per capita consumption and per capita GDP, although it is not statistically significant at the traditional levels.

Table 3: Per capita consumption and per capita GDP for all provinces or cities in the People's Republic of China, 2008 (values are in yuan of RMB).

Area	Y	X	Area	Y	X
Beijing	18911	58204	Hubei	6513	16206
Tianjin	11957	46122	Hunan	6240	14492
Hebei	5674	19877	Guangdong	12663	33151
Shānxi	5525	16945	Guangxi	4987	12555
Neimenggu	7062	25393	Hainan	5552	14555
Liaoning	7965	25792	Chongqing	6545	14660
Jilin	6675	19382	Sichuan	5259	12893
Heilongjiang	5986	18478	Guizhou	4057	6915
Shanghai	24260	66367	Yunnan	4553	10540
Jiangsu	9659	33928	Xizang	3215	12109
Zhejiang	12569	37411	Shānxi	5272	14607
Anhui	5278	12045	Gansu	4274	10346
Fujian	8772	25908	Qinghai	4978	14257
Jiangxi	4702	12633	Ningxia	5816	14649
Shandong	8075	27807	Xinjiang	4890	16999
Henan	5141	16012			
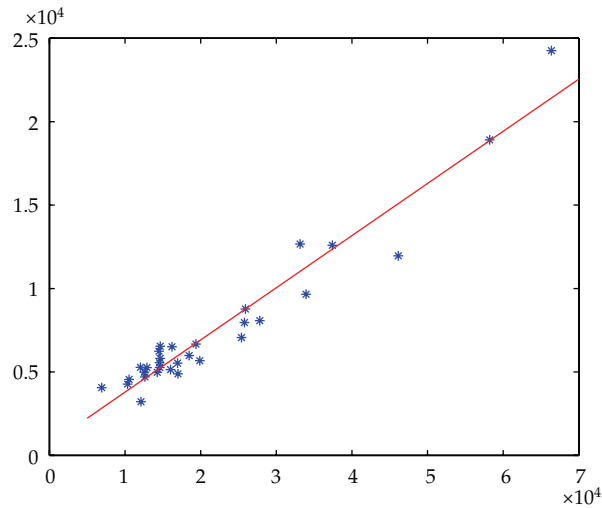
Figure 5: Regression plot of OLS.

To testify if the regression (5.1) suffers from heteroscedasticity, we obtain the residuals of the model and plotted them against with per capita GDP, as shown in Figure 6, in which we can see that the residuals are increasing around horizontal axis with the increase in per capita GDP. Therefore, there is a heteroscedasticity in the regression (5.1). Heteroscedasticity can be also obtained through other tests such as Park test and White test. So it can be said that the regression (5.1) suffers from heteroscedasticity. Furthermore, the generalized least squares (GLS) estimations $\hat{B}_0 = 665.84$ and $\hat{B}_1 = 0.31265$ can easily be got after 10000 replicates,
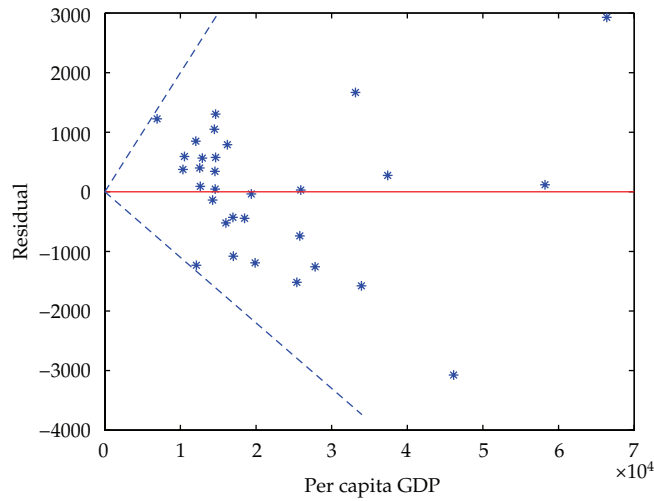
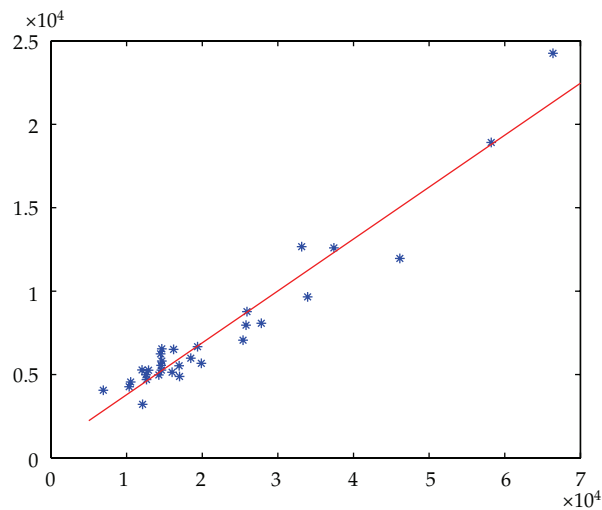**Figure 6:** Residual plot.

Figure 7: Regression plot of GLS.

provided with the bandwidth $h = 0.5$, the order $p = 2$, and MISE $= 0.275746$. Then the regression plot can be drawn in Figure 7. Compared with Figures 5 and 7, it can be said that the distribution of points around regression line in Figure 7 is more uniform than that in Figure 5. Finally, it can be exactly said that more accurate regression can be obtained by GLS than that by OLS.

6. Conclusions

In this paper we presented a new method for estimation of linear heteroscedastic regression model based on local polynomial estimation with nonparametric technique. The proposed

scheme firstly adopted the local polynomial fitting to estimate heteroscedastic function, then the coefficients of regression model are obtained based on generalized least squares method. Our approach avoided the test of heteroscedasticity for the linear model. Due to nonparametric technique of local polynomial estimation, if the heteroscedastic function is unknown, the precision of estimation was improved. Furthermore, the effect of parameters on the fitting was researched and the optimal fitting was obtained. Besides, the asymptotic normality of parameters was verified by the results of numerical simulations. Finally, the simulation results under different parameters and local polynomial estimation of real data in a case of economics really indicated that our approach was effective in finite-sample situations, which did not need to assume the form of heteroscedastic function. The presented algorithm could be easily used to heteroscedastic regression model in some practical problems.

Appendix

Suppose $X_1, \ldots, X_T$ is a stationary sequence. Let F_i^k be the σ-algebra of events generated by the random variables $\{X_j, i \leq j \leq k\}$, and $L^2(F_i^k)$ consists of F_i^k-measurable random variables with finite second moment. Before giving the Lemma A.1, we show the condition as follows.

Condition:

(i) The kernel $K(\cdot)$ is bounded with a bounded support.

(ii) $\sigma^2(\cdot)$ and $f(\cdot)$ are continuous at the points x and $f(x) > 0$.

(iii) The sequence is ρ-mixing processes, that is,

$$\rho(n) = \sup_{X \in L^2(F_{-\infty}^0), \; Y \in L^2(F_T^\infty)} |\mathrm{Corr}(X, Y)| \longrightarrow 0 \tag{A.1}$$

as $T \to \infty$. we also assume that $\Sigma_l \rho(l) < \infty$, $EY_0^2 < \infty$.

(iv) For ρ-mixing processes, there exists a sequence of positive integers satisfying $s_T \to \infty$ and $s_T = o(\sqrt{Th})$ such that $\sqrt{T/h}\rho(s_T) \to 0$ as $T \to \infty$.

Then we have the following lemma.

Lemma A.1. *Under Condition 1, if $h = O(T^{1/(2p+3)})$ and $m^{(p+1)}(\cdot)$ is continuous at the point x, then as $T \to \infty$,*

$$\sqrt{Th}\left[\mathrm{diag}(1, \ldots, h^p)\left\{\hat{\beta}(x) - \beta_0(x)\right\} - \frac{h^{p+1}m^{(p+1)}(x)}{(p+1)!}S^{-1}c_p\right] \xrightarrow{D} N\left(0, \frac{\sigma^2(x)S^{-1}S^*S^{-1}}{f(x)}\right), \tag{A.2}$$

where $\beta_0(x) = (m(x), m'(x) \ldots, m^{(p)}(x)/p!)^T$, $S = (\mu_{i+j-2})_{(p+1)\times(p+1)}$, $S^* = (\nu_{i+j-2})_{(p+1)\times(p+1)}$, and $c_p = (\mu_{p+2-j})_{(p+1)\times 1}$, and where μ_i and ν_i are defined as $\mu_i = \int_{-\infty}^{\infty} t^i K^2(t)dt$ and $\nu_i = \int_{-\infty}^{\infty} t^i K(t)dt$, respectively.

The proof of Lemma A.1 is shown in [23]. An immediate consequence of Lemma A.1 is that the asymptotic bias and the asymptotic variance for the local polynomial estimator are defined as

$$AB(x) = e_1^T S^{-1} c_p \frac{m^{p+1}(x)}{(p+1)!} h^{p+1},$$

$$AV(x) = e_1^T S^{-1} S^* S^{-1} e_1 \frac{\sigma^2(x)}{Thf(x)}. \tag{A.3}$$

Then, the mean square error (MSE) at point x is

$$\text{MSE}(x) = [AB(x)]^2 + AV(x). \tag{A.4}$$

Lemma A.2. *For the model (3.4), ε is assumed to be a vector and follows normal distribution with mean zero and covariance Σ, where $\Sigma > 0$, and $P\lim_{n\to\infty} (X'\Sigma^{-1}X/n) = Q$, a finite positive definite matrix. Then one has*

(1) *$\widehat{\beta}$ is a consistent estimator of β.*

(2) *$\widehat{\beta}$ is asymptotically normal with the mean β and covariance matrix $(X'\Sigma^{-1}X)^{-1}$.*

The proof is shown in [32].

Lemma A.3. *Suppose $\widehat{\sigma}^2(x)$ is a p-order local polynomial estimator (LPE) of the variance function $\sigma^2(x)$. Under conditions $h_n \to 0$, and $nh_n \to \infty$:*

(1) *if p is an odd, and the variance function $\sigma^2(x)$ is continuous $p+1$ derivative and bounded, then*

$$\text{MSE}\left(\widehat{\sigma}^2(x)\right) = K_1 h_n^{2p+2} \left[\frac{d^{p+1}\left(\sigma^2(x)\right)}{dx^{p+1}}\right]^2 + \frac{K_2}{nh_n} + o\left(h_n^{2p+2} + \frac{1}{nh_n}\right), \tag{A.5}$$

where K_1 and K_2 are constants. Therefore,

$$\max_{1\le i\le n} \text{MSE}\left(\widehat{\sigma}_i^2\right) = O\left(h_n^{2p+2} + \frac{1}{nh_n}\right), \tag{A.6}$$

(2) *if p is an even, and the variance function $\sigma^2(x)$ is continuous $p+1$ and $p+2$ derivative and bounded, then*

$$\text{MSE}\left(\widehat{\sigma}^2(x)\right) = h_n^{2p+4}\left[\frac{K_1 d^{p+1}\left(\sigma^2(x)\right)}{dx^{p+1}} + \frac{K_2 d^{p+2}\left(\sigma^2(x)\right)}{dx^{p+1}}\right.$$
$$\left. + \frac{K_3 d^{p+1}\left(\sigma^2(x)\right)}{dx^{p+1}} \cdot \frac{d^{p+2}\left(\sigma^2(x)\right)}{dx^{p+1}}\right] + \frac{K_4}{nh_n} + o\left(h_n^{2p+4} + \frac{1}{nh_n}\right), \tag{A.7}$$

where K_1, K_2, K_3, and K_4 are constants. Then we can get $\max_{1\le i\le n} \text{MSE}(\widehat{\sigma}_i^2) = O(h_n^{2p+4} + 1/nh_n)$.

The proof is shown in [18].

Theorem A.4. *For the model (3.4), ε is assumed to be a vector and follows normal distribution with mean zero and covariance Σ, where $\Sigma > 0$, and $p \lim_{n \to \infty} (X^T X/n) = Q_1$, $p \lim_{n \to \infty} X^T \varepsilon/\sqrt{n} = Q_2$, where Q_1 and Q_2 are positive definite matrixes. Besides, they satisfy the conditions of Lemma A.3, then β_{LPE} is asymptotically normal with mean β and covariance $(X^T \Sigma^{-1} X)^{-1}$.*

Proof. By Lemma A.2 and the bibliography [32], in order to prove the above results, we only need to testify

$$p \lim_{n \to \infty} X^T \left| \widehat{\Sigma} - \Sigma \right| X = 0, \tag{A.8}$$

where $\widehat{\Sigma}$ is an estimator of Σ, that is to testify

$$p \lim_{n \to \infty} \left| \widehat{\Sigma} - \Sigma \right| = 0. \tag{A.9}$$

By Lemma A.3, whether p is an odd or an even, we have

$$p \lim_{n \to \infty} \max \left| \widehat{\sigma}_i^2 - \sigma_i^2 \right| = 0. \tag{A.10}$$

Therefore,

$$p \lim_{n \to \infty} X^T \left| \widehat{\Sigma} - \Sigma \right| X = 0. \tag{A.11}$$

The conclusions follow. $\square$

Acknowledgments

References

[1] T. Amemiya, "A note on a heteroscedastic model," *Journal of Econometrics*, vol. 6, no. 3, pp. 365–370, 1977.

[2] P. Chen and D. Suter, "A bilinear approach to the parameter estimation of a general heteroscedastic linear system, with application to conic fitting," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 3, pp. 191–208, 2007.

[3] J. Ma and L. Liu, "Multivariate nonlinear analysis and prediction of Shanghai stock market," *Discrete Dynamics in Nature and Society*, vol. 2008, Article ID 526734, 8 pages, 2008.

[4] H. Liu, Z. Zhang, and Q. Zhao, "The volatility of the index of shanghai stock market research based on ARCH and its extended forms," *Discrete Dynamics in Nature and Society*, vol. 2009, Article ID 743685, 9 pages, 2009.

[5] H. Hu, "QML estimators in linear regression models with functional coefficient autoregressive processes," *Mathematical Problems in Engineering*, vol. 2010, Article ID 956907, 30 pages, 2010.

[6] C. R. Rao, Y. Wu, and Q. Shao, "An *M*-estimation-based procedure for determining the number of regression models in regression clustering," *Journal of Applied Mathematics and Decision Sciences*, vol. 2007, Article ID 37475, 15 pages, 2007.

[7] Q. X. He and M. Zheng, "Local polynomial regression for heteroscedaticity in the simple linear model," *Systems Engineering-Theory Methodology Applications*, vol. 12, no. 2, pp. 153–156, 2003.

[8] W. Härdle, M. Müller, S. Sperlich, and A. Werwatz, *Nonparametric and Semiparametric Models*, Springer Series in Statistics, Springer, New York, NY, USA, 2004.

[9] L. Galtchouk and S. Pergamenshchikov, "Sharp non-asymptotic oracle inequalities for non-parametric heteroscedastic regression models," *Journal of Nonparametric Statistics*, vol. 21, no. 1, pp. 1–18, 2009.

[10] J. Fan, "Local linear regression smoothers and their minimax efficiencies," *The Annals of Statistics*, vol. 21, no. 1, pp. 196–216, 1993.

[11] J. Fan and W. Zhang, "Statistical estimation in varying coefficient models," *The Annals of Statistics*, vol. 27, no. 5, pp. 1491–1518, 1999.

[12] I. Miranda Cabrera, H. L. Baños Díaz, and M. L. Martínez, "Dynamical system and nonlinear regression for estimate host-parasitoid relationship," *Journal of Applied Mathematics*, vol. 2010, Article ID 851037, 10 pages, 2010.

[13] A. Bilbao-Guillerna, M. De la Sen, A. Ibeas, and S. Alonso-Quesada, "Robustly stable multiestimation scheme for adaptive control and identification with model reduction issues," *Discrete Dynamics in Nature and Society*, vol. 2005, no. 1, pp. 31–67, 2005.

[14] T. J. Koo, "Stable model reference adaptive fuzzy control of a class of nonlinear systems," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 624–636, 2001.

[15] M. De la Sen, "Identification of a class of linear time-varying continuous-time systems," *Automation and Remote Control*, vol. 63, no. 2, pp. 246–261, 2002.

[16] M. Jankovic, "Adaptive nonlinear output feedback tracking with a partial high-gain observer and backstepping," *Transactions on Automatic Control*, vol. 42, no. 1, pp. 106–113, 1997.

[17] M. Jansson and B. Wahlberg, "On consistency of subspace methods for system identification," *Automatica*, vol. 34, no. 12, pp. 1507–1519, 1998.

[18] J. Fan, "Design-adaptive nonparametric regression," *Journal of the American Statistical Association*, vol. 87, no. 420, pp. 998–1004, 1992.

[19] M. Pandya, K. Bhatt, and P. Andharia, "Bayes estimation of two-phase linear regression model," *International Journal of Quality, Statistics, and Reliability*, vol. 2011, Article ID 357814, 9 pages, 2011.

[20] L.-Y. Su, "Prediction of multivariate chaotic time series with local polynomial fitting," *Computers & Mathematics with Applications*, vol. 59, no. 2, pp. 737–744, 2010.

[21] L. Y. Su, "Multivariate local polynomial regression with application to Shenzhen component index," *Discrete Dynamics in Nature and Society*, vol. 2011, Article ID 930958, 11 pages, 2011.

[22] L. Y. Su, Y. J. Ma, and J. J. Li, "The application of local polynomial estimation in suppressing strong chaotic noise," *Chinese Physics B*, vol. 21, no. 2, Article ID 020508, 2012.

[23] J. Fan and Q. Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer Series in Statistics, Springer, New York, NY, USA, 2003.

[24] D. Ruppert and M. P. Wand, "Multivariate locally weighted least squares regression," *The Annals of Statistics*, vol. 22, no. 3, pp. 1346–1370, 1994.

[25] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[26] H. Takeda, S. Farsiu, and P. Milanfar, "Robust kernel regression for restoration and reconstruction of images from sparse noisy data," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '06)*, pp. 1257–1260, Atlanta, Ga, USA, October 2006.

[27] L. Firinguetti, "Ridge regression in the context of a system of seemingly unrelated regression equations," *Journal of Statistical Computation and Simulation*, vol. 56, no. 2, pp. 145–162, 1997.

[28] J. Fan and I. Gijbels, "Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation," *Journal of the Royal Statistical Society B*, vol. 57, no. 2, pp. 371–394, 1995.

[29] L. Y. Su and F. L. Li, "Deconvolution of defocused image with multivariate local polynomial regression and iterative wiener filtering in DWT domain," *Mathematical Problems in Engineering*, vol. 2010, Article ID 605241, 14 pages, 2010.

[30] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*, vol. 66 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London, UK, 1996.

[31] H. C. Rutemiller and D. A. Bowers, "Estimation in a heteroscedastic regression model," *Journal of the American Statistical Association*, vol. 63, pp. 552–557, 1968.

[32] H. G. William, *Econometric Analysis*, Macmillan, New York, NY, USA, 1993.