Wavelet Regression Estimation in Longitudinal Data Analysis

ALWELL J. OYET
and
BRAJENDRA SUTRADHAR

Department of Mathematics and Statistics,
Memorial University of Newfoundland
St. John’s, NF Canada, A1C 5S7
aoyet@math.mun.ca and bsutradh@math.mun.ca

SUMMARY

Nonparametric regression models for longitudinal data are used to analyse a wide variety of repeated continuous data collected from a large number of independent individuals. The two essential aspects of this type of model are the specification of the constant nonparametric regression function for all individuals and the specification of the longitudinal correlations. As opposed to the existing kernel approaches, in this paper, we develop a wavelet regression approach for the estimation of the nonparametric function. It is shown through a simulation study that the wavelet approach estimates the function with smaller mean squared errors in comparison to the kernel approach. Also, unlike the kernel methods, the wavelet approach does not require any bandwidth. Further, it is demonstrated that even though the data are correlated, ignoring correlation yields better estimates for the regression function by both kernel and wavelet approaches. Ignoring correlation, however, has an adverse effect in forecasting a future value. This appear to hold for both kernel and wavelet approaches.

Some key words: Nonparametric function; Longitudinal correlations; Wavelet and kernel approaches; Independence approach; Forecasting.

1. INTRODUCTION

Wavelet methods have been most widely studied in the nonparametric regression problem of estimating a function $f$ on the basis of observations $y_t$ at time point $t$, modeled as

$$y_t = f(t) + \varepsilon_t, \quad t = 1, 2, \ldots, T,$$

(1.1)
where $\varepsilon_1, \ldots, \varepsilon_t, \ldots, \varepsilon_T$ are noise. Traditionally, it is assumed that $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$. See for example, Donoho and Johnstone (1995) and Abramovich and Silverman (1998). As the independence assumption for the errors in model (1.1) does not seem to be reasonable in the time series set up, it raises the issue of estimating $f$ under suitable correlation structure for the errors. This type of nonparametric estimation with correlated errors has also been discussed by many authors. For example, we refer to Altman (1990) and Schick (1996, 1998, 1999), among others.

Nonparametric inference in the clustered longitudinal set up has also received considerable attention. In this set up, one writes a nonparametric model on the basis of observations $y_{it}$ at time point $t$ for an individual $i$, as

$$y_{it} = f_i(t) + \varepsilon_{it}, \quad i = 1, 2, \ldots, I; \ t = 1, 2, \ldots, T,$$

where $I$ is usually large ($I \to \infty$) and $T$, as opposed to the model (1.1), is usually small. That is, a small number of repeated responses $\{y_{it}, \ i = 1, 2, \ldots, I; \ t = 1, 2, \ldots, T\}$ are collected from a large number of independent individuals. For an important situation when the time dependent nonparametric functions are the same for all individuals, i.e., $f_i(t) = f(t)$ for all $i = 1, 2, \ldots, I$, some authors have estimated the nonparametric function $f(t)$ by exploiting the presence of longitudinal correlations. For example, we refer to Staniswalis and Lee (1998) and Hart and Werhly (1986), among others. Note that these models in the longitudinal set up are developed for continuous response variables. For nonparametric regression analysis in the longitudinal set up, mainly for discrete variables, we refer to Lin and Carroll (2001a, 2001b), for example.

In this paper, we deal with a special semiparametric regression model

$$y_{it} = f(t) + \varepsilon_{it}, \quad (1.3)$$

where $f(t)$ is a common nonparametric function for all $i = 1, 2, \ldots, I$. As far as the error terms are concerned, the $T \times 1$ error vector $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{it}, \ldots, \varepsilon_{iT})^T$ is assumed to have a Gaussian type autocorrelation structure $V = \sigma^2C$, say, where $C$ is the common $T \times T$ autocorrelation matrix given by

$$C = \begin{pmatrix} \rho_{|t-u|} \end{pmatrix}, \quad (1.4)$$
with $\rho_{|t-u|}$ as the $|t-u|$-th lag autocorrelation between $\varepsilon_{it}$ and $\varepsilon_{iu}$, for $t \neq u$, $t, u = 1, \ldots, T$, and for $t = u$, $\rho_0 = 1$. For example, for a scalar $\rho$ satisfying $|\rho| \leq 1$, $\rho_{|t-u|} = \rho^{|t-u|}$ under the autoregressive order 1 (AR(1)) correlation structure and $\rho_{|t-u|} = \rho$ under the exchangeable correlation structure. Similarly, under moving average order 1 (MA(1)) correlation structure, one writes $\rho_{|t-u|} = \rho$ for $|t-u| = 1$ and $\rho_{|t-u|} = 0$, otherwise. To fit the model (1.3), i.e., to estimate the nonparametric function $f(t)$, Staniswalis and Lee (1998) (see also Hart and Werhly (1986)) have used a kernel approach, where the function $f$ is fitted based on a truncated generalized Fourier series using estimation of the normalized eigenfunctions of the $V$ matrix as elements of basis. These authors estimated this covariance matrix $V$ (see Staniswalis and Lee (1998, §6)) by using a weighted sample covariance matrix, where the weights are bounded corrected kernels (Rice 1984) for estimating the value of a nonparametric regression function at time point $t$ with bandwidth $h$, say. Hart and Werhly (1986, equation (2.5)) estimated the nonparametric function $f$ by using a Gasser and Müller (1979) type kernel estimator with bandwidth $h$. The selection of the bandwidth $h$ is however not easy. Staniswalis and Lee (1998) following Nadaraya (1989, p. 122) chose the bandwidth $h$ under the “working” independence assumption. Hart and Werhly (1986) chose the bandwidth $h$ that minimizes an estimated mean squared error, which is a function of the correlations of the errors.

Unlike the existing kernel approaches, we use a wavelet approach to estimate the nonparametric function $f$, after taking the longitudinal correlation into account. The specific plan of the paper is as follows. In §2, we discuss the clustered longitudinal model and the wavelet estimation approach. The performances of the proposed wavelet and existing kernel approaches, in estimating the function $f$, are compared in §3 through a simulation study. Note that both of these estimation approaches accommodate the longitudinal correlations in estimating the nonparametric function $f$. The effect of ignoring the correlation structure on the estimation of $f$ is also examined. Furthermore, in §4, we consider a one-step ahead forecasting problem and compare the forecasting performances of the wavelet and kernel approaches when they are constructed either by accommodating or ignoring the correlation structure. Some concluding remarks are given in §5.

2. LONGITUDINAL MODEL AND WAVELET ESTIMATION

Recall from (1.3) that

$$y_{it} = f(t) + \varepsilon_{it},$$

(2.1)
where \( \varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{it}, \ldots, \varepsilon_{iT})' \sim (0, \sigma^2 C) \) with \( C \) as the \( T \times T \) Gaussian type correlation matrix. To estimate the nonparametric function \( f \) by wavelet approach, we first express \( f(t) \) in terms of its finite order \( m \) wavelet expansion and rewrite the model (2.1) as

\[
y_{it} = q_m(t)\beta + \varepsilon_{it}, \quad i = 1, 2, \ldots, I; \quad t = 1, 2, \ldots, T, 
\]

(2.2)

for \( t \in [A_{i,t-1}, A_{i,t}] \) so that \( \bigcup_{i=1}^{T}[A_{i,t-1}, A_{i,t}] = [0, 1] \), for all \( i = 1, 2, \ldots, I \). In (2.2), \( q_m(t) \) is a \( 2^{m+1} \times 1 \) vector consisting of a system of dilated and translated versions \( \psi^{-l,k}(t) = 2^{l/2}\psi(2^lt - k) \), \( (l = 0, \ldots, m; \ k = 0, \ldots, 2^l - 1) \) of a primary wavelet \( \psi(t) \) and a scaling function \( \phi(t) \). For example, the Haar wavelet is defined by \( \phi(t) = I_{(0,1)}(t) \) and the primary wavelet is \( \psi(t) = \phi(2t) - \phi(2t-1) \). In (2.2), \( \beta \) is a \( 2^{m+1} \)-dimensional “working” regression vector.

Next, for a given \( t \) by averaging over the independent individuals, one obtains from (2.2) that

\[
\bar{y}_t = q_m'(t)\beta + \bar{\varepsilon}_t, \quad t = 1, 2, \ldots, T, 
\]

(2.3)

where \( \bar{y}_t = \sum_{i=1}^{I} y_{it}/I \), and \( \bar{\varepsilon}_t = \sum_{i=1}^{I} \varepsilon_{it}/I \). Let \( \bar{y} = (\bar{y}_1, \ldots, \bar{y}_t, \ldots, \bar{y}_T)' \) and \( \bar{\varepsilon} = (\bar{\varepsilon}_1, \ldots, \bar{\varepsilon}_t, \ldots, \bar{\varepsilon}_T)' \). It then follows from (2.3) that

\[
\bar{y} = Q\beta + \bar{\varepsilon}, 
\]

(2.4)

where \( Q' = [q_m(1), \ldots, q_m(t), \ldots, q_m(T)] \) is the \( 2^{m+1} \times T \) matrix with \( q_m(t) \) as defined in (2.2). We now assume that the responses of any two individuals are independent irrespective of the times of these observations. It then follows that that \( \text{cov}(\varepsilon_{it}, \varepsilon_{i'u}) = 0 \), for \( i \neq i' \), \( i, i' = 1, \ldots, I \). Consequently, \( \text{cov}(\bar{\varepsilon}) = \sigma^2 C/I = \Sigma \), say.

### 2.1. Wavelet Estimation of \( f(t) \)

Note that for the estimation of the nonparametric function \( f \), we do not require to know \( \sigma^2 \). We therefore propose to estimate \( f \) and the correlation matrix \( C \), iteratively. For known correlation matrix, one may obtain from (2.4) that

\[
z = Q'\beta + \varepsilon^*, 
\]

(2.5)
where \( z = C^{-\frac{1}{2}} \bar{y} \), \( Q^* = C^{-\frac{1}{2}} Q \), and \( \varepsilon^* = C^{-\frac{1}{2}} \bar{\varepsilon} \). It then follows that the Gasser-Müller estimator of the \( 2^{m+1} \)-dimensional vector \( \beta \), obtained from the responses of the \( i \)-th \((i = 1, 2, \ldots, I)\) individual, \( \hat{\beta}_{i,GM} \), say, is given by

\[
\hat{\beta}_{i,GM} = \sum_{t=1}^{T} z_t \int_{A_{i,t-1}}^{A_{i,t}} q^*_m(s) ds,
\]

(Gasser and Müller (1979)), where \( t \in [A_{i,t-1}, A_{i,t}] \) as in (2.2) and \( z_t \) is the \( t \)-th \((t = 1, 2, \ldots, T)\) element of the vector \( z = C^{-\frac{1}{2}} \bar{y} = (z_1, \ldots, z_t, \ldots, z_T)' \). Therefore, the Gasser-Müller estimator of \( \beta \) may be obtained as

\[
\hat{\beta}_{GM} = \frac{1}{I} \sum_{i=1}^{I} \sum_{t=1}^{T} z_t \int_{A_{i,t-1}}^{A_{i,t}} q^*_m(s) ds.
\]

Next, to obtain an improved estimate of \( f(t) \), one may modify \( \hat{\beta}_{GM} \) in (2.7) as

\[
\hat{\beta}_{MGM} = \frac{1}{I} \sum_{i=1}^{I} \sum_{t=1}^{T} z_t B_i^{-1} \int_{A_{i,t-1}}^{A_{i,t}} q^*_m(s) ds,
\]

where

\[
B_i = \sum_{t=1}^{T} \int_{A_{i,t-1}}^{A_{i,t}} q^*_m(s) q^*_m(t) ds.
\]

This yields the modified Gasser-Müller type estimator of \( f(t) \) as

\[
\hat{f}(t) = q'_m(t) \hat{\beta}_{MGM},
\]

for a suitable choice of \( m \).

### 2.2. Estimation of the Longitudinal Correlation Matrix

Note that the wavelet estimation of the nonparametric function \( f(t) \) by (2.9) requires the longitudinal correlation matrix \( C \) to be known. Recall from (1.4) that \( C = (\rho_l) \), where \( \rho_l \) is the \( l \)-th \((l = 1, 2, \ldots, T - 1)\) lag autocorrelation. Following the idea of pooling several independent time series (see Quenouille (1958)), Jowaheer and Stutradhar (2002) recently suggested moment estimators for the autocorrelations in a longitudinal set up for discrete data. We follow this suggestion and write the moment estimator of \( \rho_l \) as

\[
\hat{\rho}_l = \frac{\sum_{i=1}^{I} \sum_{t=1}^{T-l} (y_{it} - \hat{f}(t))(y_{i,t+l} - \hat{f}(t+l)) \left/ \{I(T-l)\} \right.}{\sum_{i=1}^{I} \sum_{t=1}^{T-l} (y_{it} - \hat{f}(t))^2 / IT},
\]

where, for \( t = 1, \ldots, T, \hat{f}(t) \) is given by (2.9). This moment estimator \( \hat{\rho}_l \) is consistent for \( \rho_l \). The estimation of \( f(t) \) by (2.9) and \( C = (\rho_l) \) by (2.10) is carried out iteratively as follows.
1. **Step 1:** We use $\hat{\rho}_l = 0$, i.e., $C_0 = I_T$ to estimate $f(t)$ by (2.9). Call this $f_0(t)$. 

2. **Step 2:** Use the estimate of $f(t)$ denoted by $f_0(t)$ in (2.10), to compute an estimate of $\rho_l$ ($l = 1, 2, \ldots, T - 1$).

3. **Step 3:** Construct the $C$ matrix with the estimate of $\rho_l$ obtained in Step 2. Call this new $C$ matrix $C_1$. Now, use $C_1$ to compute a new estimate of $f(t)$. Call this $f_1(t)$.

4. **Step 4:** Replace $f_0(t)$ by $f_1(t)$ and repeat Steps 2 and 3 to obtain $C_2$ and $f_2(t)$. Continue this cyclical iteration until convergence is achieved. The values at which convergence occurs are then the estimates for $f(t)$ and $\rho_l$.

3. **WAVELET VERSUS KERNEL ESTIMATION: A SIMULATION STUDY**

Two nonlinear functions were used to investigate and compare the performance of the wavelet and kernel estimators. These functions include (1) a nonlinear function,

$$f_1(t) = \frac{\phi_1 \sin \left( \frac{t}{T + 0.5} \right)}{\phi_2 + \sin \left( \frac{t}{T + 0.5} \right)} + \phi_3 \sin \left( \frac{t}{T + 0.5} \right),$$

which has been used by Lindstrom and Bates (1990) (see also Das and Sutradhar (1995)), among others, to model the so-called guinea pig data, and (2) the HeavySine function defined by

$$f_2(t) = 4 \sin \left( \frac{4\pi t}{T + 0.5} \right) - \text{sgn} \left( \frac{t}{T + 0.5} - 0.2 \right) - \text{sgn} \left( 0.72 - \frac{t}{T + 0.5} \right),$$

for $t = 1, \ldots, T$.

We have considered $I = 100$ and $T = 6, 10$ and 18, and generated a data set of $T$ points for each of the 100 individuals by using the nonlinear function $f_1(t)$, ($t = 1, \ldots, T$) with $\phi_1 = 0.3$, $\phi_2 = 0.5$ and $\phi_3 = 0.01$. As far as the errors of the model (2.1) are concerned, we have generated $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ following an $AR(1)$ process $\varepsilon_t = \rho \varepsilon_{i,t-1} + a_t$; $a_t \overset{iid}{\sim} N(0, \sigma^2)$, with $\sigma^2 = 1$ and $\rho = 0.5, 0.7$ and 0.9. To generate data by using the HeaviSine function, we have considered the same $I = 100$ and $T = 6, 10$ and 18. Also, the errors were generated following the $AR(1)$ process similar to that for the $f_1(t)$ function. For the purpose of illustration, we exhibit plots of data points generated by using $f_1(t)$ and the fitted curve when $T = 10$, $\rho = 0.7$ and data points generated by using $f_2(t)$ and the fitted curve when $T = 6$, $\rho = 0.9$. 

[Insert Figure 1 about here]
For the wavelet estimation of the functions $f_1(t)$ and $f_2(t)$ we have used $m = 1, 2$ and 3 for the cases with $T = 6, 10$ and 18, respectively, $m$ being the order of wavelet expansion. More specifically, for a selected value of $m$, we have used the steps outlined in §2.2 to iteratively estimate $f(t)$ and $\rho_l$ (lag $l$ autocorrelation). The whole estimation procedure is repeated 1000 times. Table 1 shows the mean squared error (MSE) of the wavelet estimator computed from these 1000 simulations, for selected values of $\rho$. To be specific, the MSE was computed by

$$MSE\left(\hat{f}(t)\right) = \frac{1}{1000} \sum_{s=1}^{1000} (\hat{f}^{(s)}(t) - f(t))^2,$$

where $\hat{f}^{(s)}(t)$ denotes the wavelet estimate of $f(t)$ obtained at the $s$-th ($s = 1, \ldots, 1000$) simulation. Note that the wavelet estimator of $f(t)$ by (2.9) is referred to as the modified Gasser-Müller (MGM) estimator. We have also estimated the nonlinear function by using the well-known weighted least squares (WLS) method. The weights used are those derived by Oyet and Wiens (2000). The MSE for the WLS approach in estimating $f(t)$ based on the wavelet technique is also shown in Table 1.

Furthermore, we have computed the MGM and WLS estimators for a selected function under the so-called “working” independence assumption. To be specific, even though the data $\{y_{it}\}$ were generated by following the AR(1) error structure with correlation parameters $\rho_l$ ($l = 1, \ldots, T − 1$), in our estimation, we have however, used $\hat{\rho}_l = 0$ for all $l = 1, \ldots, T − 1$. The longitudinal independence assumption based MGM and WLS estimators are denoted by MGM(IND) and WLS(IND) respectively. The simulation based MSE of these estimators are also shown in Table 1.

In the same table we also provide the MSE of a kernel estimator computed by first exploiting the correlation structure, and then by using the “working” independence approach. These two kernel based Gasser-Müller estimators are referred to as the KGM and KGM(IND) estimators, respectively. In this kernel approach, for a given bandwidth $h$, the nonparametric estimate of the function $f(t)$ is given by

$$\tilde{f}_h(t) = \frac{1}{h} \sum_{t=1}^{T} \tilde{y}_t \int_{s_{t-1}}^{s_t} K\left(\frac{t-u}{h}\right) du,$$

(Hart and Werhly (1986)), where $\tilde{y}_t = \sum_{i=1}^{T} y_{it}/I$, $\{s_{t-1}, s_t\} = [0, 1]$, and $K$ is a density function with support $[-1, 1]$. The Epanechnikov kernel $K(y) = 0.75(1 - y^2)I_{[-1,1]}(y)$, which
was used by Hart and Werhly (1986), was also used in our simulation study. The bandwidth $h$ is chosen such that an estimated mean average squared error (MASE) curve is minimized, where the MASE may be written as

$$M(h) = \frac{1}{T} \sum_{t=1}^{T} \left[ \tilde{f}_h(t) - f(t) \right]^2,$$  \hspace{1cm} \text{(3.2)}

where $\tilde{f}_h(t) = \tilde{f}_h(t)/a_h(t)$ with $\tilde{f}_h(t)$ as in (3.1) and

$$a_h(t) = \frac{1}{h} \sum_{t=1}^{T} \int_{s_{t-1}}^{s_t} K \left( \frac{t-u}{h} \right) du.$$  

The MASE in (3.2) further may be expressed as

$$M(h) = \frac{1}{T} E \{ RSS(h) \} - \frac{\sigma^2}{T} \left[ 1 - 2T^{-1} \text{trace}(K_h C) \right],$$ \hspace{1cm} \text{(3.3)}

where $RSS(h) = \sum_{t=1}^{T} \left[ \tilde{y}_t - \tilde{f}_h(t) \right]^2$ and $K_h$ is a $T \times T$ matrix with $(t,v)$-th element $h^{-1} \int_{s_{t-1}}^{s_t} K \left( \frac{v-u}{h} \right) du/a_h(v)$. In (3.3), $\text{trace}(K_h C)$ may be simplified as

$$\text{trace}(K_h C) = 2b_{T,h}(0) \int_{0}^{1/2h} K(y) dy + \sum_{l=1}^{T-1} b_{T,h}(l) \rho(l) \int_{(l-1/2)/h}^{(l+1/2)/h} K(y) dy,$$ \hspace{1cm} \text{(3.4)}

where

$$b_{T,h}(l) = \sum_{u=1}^{T-l} a_h^{-1}(u) \quad \text{and} \quad \rho(l) = \rho_l, \text{ for } l = 1, \ldots, T - 1.$$  

Note that in the existing kernel approach, the bandwidth $h$ is determined based on the correlation matrix through (3.4), whereas in the proposed wavelet approach discussed in §2, the selection of $m$, the order of the wavelet expansion has nothing to do with the correlation structure. Further note that as $l$ gets larger, the value of the second integral in (3.4) approaches zero. It then follows that for large $l$, the lag correlations $\rho_l$ contribute very little in determining the value of $h$. In addition, for the Epanechnikov kernel $K(y) = 0.75(1 - y^2)I_{(-1,1)}(y)$ (see Hart and Werhly 1986) the second integral in (3.4) is zero when $h$ is small. Hence, when $h$ is small the MSE under the KGM approach is the same as the MSE under the KGM(IND) approach (see Table 1) and the lag correlations $\rho_l$ does not contribute to $M(h)$. Consequently, since $M(h)$ is minimized at $h = 0.1$ (see Table 1) and the lag correlations $\rho_l$ did not contribute to $M(0.1)$, it is not reasonable to claim, as in Hart and Werhly (1986), that the lag correlations contribute to the estimation of $f(t)$ once we chose $h$ on the basis of $M(h)$.
The results of Table 1 show that between the two wavelet approaches, the proposed MGM estimators always yield the same or smaller MSE when compared to the WLS estimators in estimating the function $f(t)$ nonparametrically. For example, for $T = 10$ the MGM estimator has MSE = 0.025 in estimating $f_1(t)$ for the case with $\rho = 0.9$, whereas WLS yields MSE = 0.107. When MGM(IND) and WLS(IND) are compared, they appear to yield almost the same MSE in estimating $f_1(t)$. This behaviour of the MGM and WLS estimators also appear to hold for the estimation of the HeaviSine function. For example, except for $T = 10$ and $\rho = 0.9$ case, the MGM estimator appear to yield smaller MSE than the WLS estimator for all other values of $T$ and $\rho$. Once again the MSE of the MGM(IND) and WLS(IND) estimators are the same.

As the MGM approach is better than the WLS approach, we next compare the performance of this MGM approach with the existing kernel estimator KGM. In the kernel approach, the $f_1(t)$ and $f_2(t)$ functions were estimated by (3.1) where $h$ was selected by minimizing the MASE, i.e., $M(h)$, given in (3.3). We computed $M(h)$ for values of $h$ ranging from 0.01 to 5.00 at intervals of 0.01. It was found that in almost all cases, the minimizing $h$ was close to 0.1. We, therefore choose to report the MSE of the KGM and KGM(IND) for $h = 0.10$. It is found from Table 1 that except for the estimation of $f_1(t)$ for $T = 6$ case, the wavelet based MGM estimator almost always yields a smaller MSE as compared to the kernel based KGM estimator. For example, in estimating the HeaviSine function $f_2(t)$ for $T = 6$, the MGM approach yields uniformly smaller MSE for all $\rho = 0.5, 0.7$ and 0.9, as compared to the KGM estimator. Similarly, the MGM(IND) estimator appear to perform better than the KGM(IND) estimator in all cases. Thus, in general, the proposed wavelet approach performs better than the existing kernel approaches in all longitudinal set up, whether $T$ is small or large.

Note however that the “working” independence assumption based estimation appear to perform better than the correlation structure based estimators both in the wavelet and kernel approaches. However, when $T$ is large, the MSE of the “working” independence assumption based estimators and the MSE of the correlation structure based estimators both in the wavelet and kernel approaches are about the same. See results for $(T, m) = (18, 3)$ in Table 1. Overall, MGM(IND) appear to provide smaller MSE than the MGM approach. Similarly, KGM(IND) estimator has the smaller MSE than the KGM estimator. This good behaviour of
the independence assumption based estimators was also reported in Hart and Werhly (1986), among others, for the kernel approach. This raises the question about the necessity of the correlation based estimation either by using the wavelet or kernel approach. In the next section, we however show that the exploitation of the longitudinal correlation structure is quite important for example in forecasting a future response which is an important statistical problem.

4. WAVELET APPROACH IN FORECASTING

In addition to the estimation of the nonparametric function, the estimation of the longitudinal correlations may also be of direct interest. To illustrate this need, suppose that the errors of the semiparametric model (2.1) follow an AR(1) process, so that

\[ \varepsilon_{it} = \rho \varepsilon_{i,t-1} + a_{it}, \quad (4.1) \]

where \( a_{it} \overset{iid}{\sim} N(0, \sigma^2) \). Further suppose that we are interested in making a one-step ahead forecast under the present longitudinal set up. Following traditional time series approach, one may forecast the future value of \( y_{i,T+1} \), the response to occur at time point \( T + 1 \) for the \( i \)-th (\( i = 1, \ldots, I \)) individual, by using the forecasting equation

\[ \hat{y}_{i,T+1} = \hat{f}(T + 1) + \hat{\rho}_1 \left( y_{iT} - \hat{f}(t) \right), \quad (4.2) \]

where \( \rho_1 \) is the estimate of the lag 1 autocorrelation which may be obtained by (2.10). Note that in the “working” independence approach, one uses

\[ \hat{y}_{i,T+1} = \hat{f}(T + 1). \quad (4.3) \]

As far as the estimation of the function \( f \) at time point \( T + 1 \) is concerned, one may obtain this estimate by extrapolation based on the formulas (2.9) and (3.1) under the wavelet and kernel approaches, respectively.

In order to compare the forecasting performance of the independence approach with that of the correlation based approach, we compute the mean squared errors for the estimators in (4.3) and (4.2) based on 1000 simulations. That is, we compute the forecasting MSE (FMSE, say) as

\[ \text{FMSE} = \frac{1}{1000} \sum_{s=1}^{1000} \sum_{i=1}^{I} \left( y^{(s)}_{i,T+1} - \hat{y}^{(s)}_{i,T+1} \right)^2 / \{1000I\}, \quad (4.4) \]
for the correlation model based approach and by

\[ FMSE(IND) = \sum_{s=1}^{1000} \sum_{i=1}^{I} \left( y_{i,T+1}^{(s)} - \hat{y}_{i,T+1}^{(s)} \right)^2 / {1000I}, \]  

(4.5)

for the independence assumption based forecasting approach. In (4.4) and (4.5), \( y_{i,T+1}^{(s)} \) are, however, generated in the \( s \)-th simulation by following the AR(1) correlation structure (4.1) based semiparametric model (2.1). The FMSE and FMSE(IND) for both wavelet and kernel approaches are reported in Table 2. We also exhibit the average of the future observations \( y_{i,T+1} \) and their corresponding forecasted values \( \hat{y}_{i,T+1} \) and \( \hat{\hat{y}}_{i,T+1} \) in Figure 2, for all \( i = 1, 2, \ldots, I \) and \( T = 10 \).

Let \( x = (t^* - u)/h \), where \( t^* = t/(T + 0.5) \) and \( 1/(T + 0.5) \leq u \leq T/(T + 0.5) \). Then, \( x \) is larger than 1 for all values of \( u \), when \( T = 6 \) and \( h = 0.1 \). It follows that, in this case, \( K(x) = 0 \) under the Epanechnikov kernel. This implies that the forecast \( \hat{\hat{y}}_{i,T+1} \) cannot be computed when \( T = 6 \) if the kernel approach is used. We have therefore entered NA in the column for FMSE, in Table 2, for this case.

It is clear that the forecasted values based on the independence assumption are very different from the actual as compared to the forecasted values when the longitudinal correlation are taken into account. Furthermore, it is clear from Table 2 that the values of FMSE are always smaller than that of FMSE(IND) both under wavelet and kernel approaches. This shows that the correlation structure plays an important role in forecasting a future response. When the proposed wavelet approach is compared to the kernel approach, the former performs uniformly better in forecasting the future one-step ahead response.

In practice, the bandwidth that was used for estimation is the bandwidth that one will also use for forecasting. However, in our simulation study on forecasting we found that, for the KGM approach, the bandwidth \( h \) at which the MSE and FMSE were minimized are quite distinct; whereas, the same value of \( m \) will be used for both estimation and forecasting in the wavelet approach. This is therefore a serious drawback of the KGM approach. Thus, in analyzing longitudinal data for both estimation and forecasting, the wavelet approach is recommended.
5. CONCLUDING REMARKS

The results of this paper suggest that when estimating the common mean response curve in a longitudinal set up, it is best to use the “working” independence assumption in both wavelet and kernel methods. However, if it is of interest to forecast future observations, the longitudinal correlation structure should be exploited in forecasting, in both wavelet and kernel approaches. When wavelet and kernel methods are compared, the wavelet approach appear to outperform the kernel approach both in estimation and forecasting. This is because the MSE of the wavelet estimation were always found to be smaller than that of the kernel method. Furthermore, it was observed that the “optimal” bandwidth for estimation was different from the “optimal” bandwidth for forecasting under the kernel approach, whereas, no such bandwidth is needed in the wavelet approach. It is therefore advisable to avoid the kernel approach if the analysis requires both estimation and forecasting.

REFERENCES


Table 1. Mean Squared Error of Estimated Mean Response Functions, $\hat{f}_1(t)$ and $\hat{f}_2(t)$ Based on MGM, WLS and KGM Approaches

<table>
<thead>
<tr>
<th>$(T, m)$</th>
<th>$\rho$</th>
<th>MSE of Estimated Mean Response $\hat{f}_1(t)$</th>
<th>MSE of Estimated Mean Response $\hat{f}_2(t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Wavelet Approach</td>
<td>Kernel Approach</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MGM</td>
<td>MGM(IND)</td>
</tr>
<tr>
<td>0.9</td>
<td>0.1386</td>
<td>0.026</td>
<td>0.1881</td>
</tr>
<tr>
<td>(6,1)</td>
<td>0.7</td>
<td>0.055</td>
<td>0.0252</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.032</td>
<td>0.0246</td>
</tr>
<tr>
<td>0.9</td>
<td>0.0252</td>
<td>0.0122</td>
<td>0.1065</td>
</tr>
<tr>
<td>(10,2)</td>
<td>0.7</td>
<td>0.0134</td>
<td>0.0121</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0122</td>
<td>0.0119</td>
</tr>
<tr>
<td>0.9</td>
<td>0.0093</td>
<td>0.0093</td>
<td>0.0118</td>
</tr>
<tr>
<td>(18,3)</td>
<td>0.7</td>
<td>0.0095</td>
<td>0.0095</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0095</td>
<td>0.0095</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$(T, m)$</th>
<th>$\rho$</th>
<th>Wavelet Approach</th>
<th>Kernel Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MGM</td>
<td>MGM(IND)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1069</td>
<td>0.1069</td>
</tr>
<tr>
<td>0.9</td>
<td>0.1333</td>
<td>0.1173</td>
<td>0.1063</td>
</tr>
<tr>
<td>(6,1)</td>
<td>0.5</td>
<td>0.1117</td>
<td>0.1057</td>
</tr>
<tr>
<td>0.9</td>
<td>1.0168</td>
<td>0.09339</td>
<td>3.395</td>
</tr>
<tr>
<td>(10,2)</td>
<td>0.7</td>
<td>0.1539</td>
<td>0.09342</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1009</td>
<td>0.09313</td>
</tr>
<tr>
<td>0.9</td>
<td>0.0361</td>
<td>0.02466</td>
<td>0.4809</td>
</tr>
<tr>
<td>(18,3)</td>
<td>0.7</td>
<td>0.0254</td>
<td>0.02482</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0249</td>
<td>0.02482</td>
</tr>
</tbody>
</table>
Table 2. One-Step Ahead Forecast Mean Squared Error
Based on MGM, WLS and KGM Approaches

Forecast MSE Based on Observations Generated from $f_1(t)$

<table>
<thead>
<tr>
<th>$(T,m)$</th>
<th>$\rho$</th>
<th>Wavelet Approach</th>
<th>Kernel Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(T,m)$</td>
<td>MGM</td>
<td>MGM(IND)</td>
<td>WLS</td>
</tr>
<tr>
<td>(6,1)</td>
<td>0.9</td>
<td>2.033</td>
<td>1.6197</td>
</tr>
<tr>
<td>0.7</td>
<td>1.4037</td>
<td>1.6184</td>
<td>1.1099</td>
</tr>
<tr>
<td>0.5</td>
<td>1.5035</td>
<td>1.619</td>
<td>1.3977</td>
</tr>
<tr>
<td>(10,2)</td>
<td>0.9</td>
<td>0.2352</td>
<td>1.0466</td>
</tr>
<tr>
<td>0.7</td>
<td>0.5574</td>
<td>1.0488</td>
<td>0.5546</td>
</tr>
<tr>
<td>0.5</td>
<td>0.8010</td>
<td>1.0542</td>
<td>0.8022</td>
</tr>
<tr>
<td>(18,3)</td>
<td>0.9</td>
<td>1.7506</td>
<td>2.5582</td>
</tr>
<tr>
<td>0.7</td>
<td>2.071</td>
<td>2.563</td>
<td>2.071</td>
</tr>
<tr>
<td>0.5</td>
<td>2.3134</td>
<td>2.567</td>
<td>2.3119</td>
</tr>
</tbody>
</table>

Forecast MSE Based on Observations Generated from $f_2(t)$

<table>
<thead>
<tr>
<th>$(T,m)$</th>
<th>$\rho$</th>
<th>Wavelet Approach</th>
<th>Kernel Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(T,m)$</td>
<td>MGM</td>
<td>MGM(IND)</td>
<td>WLS</td>
</tr>
<tr>
<td>(6,1)</td>
<td>0.9</td>
<td>9.2049</td>
<td>73.79</td>
</tr>
<tr>
<td>0.7</td>
<td>29.836</td>
<td>73.77</td>
<td>64.67</td>
</tr>
<tr>
<td>0.5</td>
<td>48.423</td>
<td>73.75</td>
<td>65.67</td>
</tr>
<tr>
<td>(10,2)</td>
<td>0.9</td>
<td>3.467</td>
<td>5.183</td>
</tr>
<tr>
<td>0.7</td>
<td>4.378</td>
<td>5.188</td>
<td>3.902</td>
</tr>
<tr>
<td>0.5</td>
<td>4.698</td>
<td>5.198</td>
<td>4.160</td>
</tr>
<tr>
<td>(18,3)</td>
<td>0.9</td>
<td>4.733</td>
<td>5.607</td>
</tr>
<tr>
<td>0.7</td>
<td>5.139</td>
<td>5.639</td>
<td>5.617</td>
</tr>
<tr>
<td>0.5</td>
<td>5.429</td>
<td>5.688</td>
<td>5.875</td>
</tr>
</tbody>
</table>
and KGM approaches, mean response curve data points

Figure 1: A time plot of (a) average of data points \( \sum_{k=1}^{1000} y_{tk}/1000 \) generated from \( f_1(t) \) over 1000 simulations (\( T = 10, \rho = 0.7 \)); (b) fitted curves based on MGM, WLS with \( m = 2 \) and KGM approaches, mean response curve \( f_1(t) \) and \( y-bar = \sum_{k=1}^{1000} \bar{y}_{tk}/1000 \); (c) average of data points \( \sum_{k=1}^{1000} y_{tk}/1000 \) generated from \( f_2(t) \) over 1000 simulations (\( T = 6, \rho = 0.9 \)); (d) fitted curves based on MGM, WLS with \( m = 1 \) and KGM approaches, mean response curve \( f_2(t) \) and \( y-bar = \sum_{k=1}^{1000} \bar{y}_{tk}/1000 \).
Figure 2: A plot of average of observations at time point $T + 1$ (solid line) (a) generated by using $f_1(t)$ ($T = 18, \rho = 0.7$) and forecasted observations based on MGM (dotted lines), WLS (dotted-broken lines) with $m = 3$ and KGM (broken lines) approaches; (b) generated by using $f_2(t)$ ($T = 10, \rho = 0.9$) and forecasted observations based on MGM (dotted lines), WLS (dotted-broken lines) with $m = 2$ and KGM (broken lines) approaches.