JIRSS (2017)
Vol. 16, No. 02, pp 33-50
DOI:10.22034/jirss.2017.16.03

Two-step Smoothing Estimation of the Time-variant Parameter with Application to Temperature Data

Mohammed Chowdhury[1], Lewis VanBrackle[1] and Mohammad Patwary[2]

[1]Department of Statistics and Analytical Sciences, KSU, Georgia, USA
[2]Department of Statistics, University of Texas at Dallas, USA

Received: 26/03/2016, Revision received: 21/02/2017, Published online: 27/11/2017

Abstract. In this article, we develop two nonparametric smoothing estimators for parameter of a time-variant parametric model. This parameter can be from any parametric family or from any parametric or semi-parametric regression model. Estimation is based on a two-step procedure, in which we first get the raw estimate of the parameter at a set of disjoint time points and then compute the final estimator at any time by smoothing the raw estimators. We will call these estimators two-step local polynomial smoothing estimator and two-step kernel smoothing estimator. We derive these two two-step smoothing estimators by modeling raw estimates of the time-variant parameter from any regression model or probability model and then establish a mathematical relationship between these two estimators. Our two-step estimation method is applied to temperature data from Dhaka, the capital city of Bangladesh. Extensive simulation studies under different cross-sectional and longitudinal frameworks have been conducted to check the finite sample MSE of our estimators. Narrower bootstrap confidence bands and smaller MSEs from application and simulation results show the superiority of the local polynomial smoothing estimator over the kernel smoothing estimator.

Keywords. Bandwidth, Kernel smoothing, Local polynomials, Raw estimates, Two-

Mohammed Chowdhury (chowdhury@kennesaw.edu)

step smoothing

MSC: 62G05; 62G08.

1 Introduction

Kernel smoothing estimation and local polynomial smoothing estimation are two popular smoothing methods in nonparametric regression. Kernel smoothing estimation was developed and used by Nadaraya-Watson (1964), and since then it has been used in many application. On the other hand, local polynomial smoothing was first studied by Stone (1977, 1980, 1982) and Cleveland (1979) and then by Fan (1992, 1993), Fan and Gijbels (1992, 1996) and Ruppert and Wand (1994) and others. These two estimators are widely used when the parametric form of the response variable is unknown. In this article, the response variable is considered as the raw estimates of the parameters from a time-variant parametric model. In the first step, raw estimates are usually obtained by classical methods such as method of maximum likelihood or method of moments. If raw estimates are maximum or minimum of a variable of interest or any other quantities, we can obtain them empirically or non-parametrically. These raw estimates of the time-variant parameters are treated as data points of the response variable. In the second step, raw estimates are smoothed by applying smoothing estimators such as local polynomial or kernel estimators as derived in Section 3.

In order to model and estimate parameters directly on all time design points, we have adopted a two-step smoothing methods. Motivated by the works of Hoover *et al.* (1998) on time-varying coeficient models, we propose in this paper a structural nonparametric approach for the parameter estimation, and we show that, when certain structural assumptions hold, our method leads to estimators at any point within the entire time design point. Our approach relies on the assumption that the conditional probability model of the variable of interest follows a parametric family, but the parameters may change with time. For the estimation method, we first obtain raw estimates of the parameters based on the time-varying parametric family at a set of distinct time points, and then compute the final estimators at any time point by smoothing the available raw estimates using a nonparametric smoothing procedure. The two-step smoothing procedure, which is similar to the ones used in Fan and Zhang (2000), Wu *et al.* (2010) and Wu and Tian (2013a,b), is computationally simple and easy to implement in practice. All statistical computations and simulations are accomplished using Statistical Software R.

In the main results, we describe time-varying parametric models in Section 2, and present our derivation of these two-step smoothing estimators in Section 3. Application of our procedures is presented in Section 4. Results from extensive simulation studies under different cross-sectional and longitudinal frameworks are shown as the ratio of MSEs from these two-step smoothing estimators in Section 5. Finally, we briefly discuss in Section 6 some further implications and extensions of the theory and applications for parameter estimation.

2 Time-Variant Parametric Models

When $H_t(\cdot)$ belongs to a parametric family at each $t \in \tau$, we have the time-varying parametric model

$$\mathcal{H}_{\eta(t)} = \{H_{t,\eta(t)}(\cdot); \eta(t) \in \Omega\}, \tag{2.1}$$

where $\eta(t)$ is the vector of time-varying parameters which belong to an open Euclidean space Ω. If $\eta(t)$ is the parameter from regression model, then the time-variant parametric regression model is defined by

$$Y(t) = X(t)^T \eta(t) + \epsilon(t). \tag{2.2}$$

Chowdhury *et al.* (2016) showed that when the time-varying parametric assumption holds, a smoothing estimator that utilizes the local parameter structures in a two-step smoothing estimation of conditional CDF with local Box-Cox transformation on response variable could be superior to the kernel smoothing estimator.

3 Two - Step Estimation

We derive here two-step smoothing methods for parameter estimation, in which we first compute the raw estimates of $\eta(t_j)$ for all $j = 1, \ldots, J$, and then derive the smoothing estimates of $\eta(t)$ for any $t \in \tau$ by applying the smoothing procedure over the corresponding raw estimates at $\boldsymbol{t}$. This two-step smoothing approach is computationally simple and does not need correlation assumptions across different time points even if the data are longitudinal. This is because the data are split by time (age) by rounding the time points in a small neighborhood and hence the chance of obtaining more than one repeated measurement of a subject at any specific time point is zero. For example, if repeated measurements of a subject are collected once a year, then the chance of having two measurements of a subject in a one month period (a specific time point) is zero. More specifically, if yearly collected longitudinal data are split by months, then there is no chance of having two repeated measurements of a subject at any specific month.

3.1 Raw Estimates

We derive the estimators $\widetilde{\eta}(t_j)$ of $\eta(t_j)$ using observations at time $t_j \in \boldsymbol{t}$. Suppose that we have enough observations n_j at t_j, so that $\eta(t_j)$, $t_j \in \boldsymbol{t}$, can be estimated by the maximum likelihood estimators (MLE) $\widetilde{\eta}(t_j)$ using the subjects in $\mathcal{S}_j$. If the response variable does not follow a parametric model to allow estimation of the parameters by MLE or method of moments, we can estimate the parameters either empirically or non-parametrically.

In practice, these raw estimators require the number of observations n_j at t_j to be sufficiently large, so they can be computed numerically. When the local sample size n_j is not sufficiently large, we can round off or group some of the adjacent time points into small bins, and compute the raw estimates within each bin. This rounding off or binning approach has been used by Fan and Zhang (2000) and Chowdhury *et al.* (2016).

3.2 Smoothing Estimators

3.2.1 Two-step Local Polynomial Smoothing

Suppose that $\eta(t)$ is (p+1) times continuously differentiable with respect to $t \in \tau$. Let $\eta^{(q)}(t)$ be the qth derivative of $\eta(t), 1 \leq q \leq p$ and $\delta_q(t) = \eta^{(q)}(t)/q!$. By the Taylor expansion of $\eta(t)$,

$$\eta(t) \approx \sum_{q=0}^{p} \delta_q(a_0)(t_j - a_0)^q,$$

for t in some neighborhood of a_0. We can treat the raw estimates $\widetilde{\eta}(t_j)$ as the "observations" of $\eta(t_j)$ at t_j, $j = 1, \ldots, J$, and obtain the pth local polynomial estimators by minimizing

$$\sum_{j=1}^{J} \left\{ \widetilde{\eta}(t_j) - \sum_{q=0}^{p} \delta_q(t)(t_j - t)^q \right\}^2 K_h(t_j - t_0),$$

where $K_h(t_j - t) = K[(t_j - t)/h]/h$, $K(\cdot)$ is a non-negative kernel function, and $h > 0$ is a bandwidth. Using the matrix formulation, we define $\widetilde{\eta}(\boldsymbol{t}) = (\widetilde{\eta}(t_1), \ldots, \widetilde{\eta}(t_J))^T$, $\delta(t) = (\delta_0(t), \ldots, \delta_p(t))^T$, $G(t;h) = \text{diag}\{K_h(t_j - t)\}$ with jth column $G_j(t;h) = (0, \ldots, K_h(t_j - t), \ldots, 0)^T$, and $T_p(t)$ the $J \times (p+1)$ matrix with its jth row given by $T_{j,p}(t) = (1, t_j - t, \ldots, (t_j - t)^p)$. The local polynomial estimators $\widehat{\delta_q}(t)$ minimize

$$Q_G\Big[\delta(t)\Big] = \Big[\widetilde{\eta}(\boldsymbol{t}) - T_p(t)\delta(t)\Big]^T G(t;h)\Big[\widetilde{\eta}(\boldsymbol{t}) - T_p(t)\delta(t)\Big].$$

The pth order local polynomial estimator of $\eta^{(q)}(t)$ based on $\widetilde{\eta}(t_j)$, which minimizes $Q_G\Big[\delta(t)\Big]$, is

$$\widehat{\eta}^{(q)}(t) = \sum_{j=1}^{J} \Big\{ W_{q,p+1}(t_j, t; h)\, \widetilde{\eta}(t_j) \Big\}, \tag{3.1}$$

where $W_{q,p+1}(t_j, t; h) = q! e_{q+1,p+1} \Big[T_p^T(t) G(t; h) T_p(t)\Big]^{-1} \Big[T_{j,p}^T(t) G_j(t; h)\Big]$ is the "equivalent kernel function" (*e.g.* Fan and Zhang , 2000) and $e_{q+1,p+1}$ is the row vector of length $p + 1$ with 1 at its $(q + 1)$th place and 0 elsewhere.

By definition of $\delta(t)$, we have $\widehat{\delta}(t) = \Big(\widehat{\delta}_0(t), \ldots, \widehat{\delta}_p(t)\Big)^T$ and $\widehat{\eta}^{(q)}(t) = \widehat{\delta}_q(t)\, q!$ is an estimator for $\eta^{(q)}(t)$, $q = 0, 1, \ldots, p$. For local polynomial fitting, $p - q$ should be taken to be odd as shown in Ruppert and Wand (1994) and Fan and Gijbels (1992, 1996). When $p = 1$, we get the local linear estimator $\widehat{\eta}_L(t) = \widehat{\delta}_0(t)$ of $\eta(t)$ based on (3.1) and the equivalent kernel function $W_{0,2}(t_j, t; h)$. So, the local linear estimator is

$$\widehat{\eta}_L(t) = \widehat{\eta^{(0)}}(t|x). \tag{3.2}$$

3.2.2 Two-step Kernel Smoothing

Suppose $f(t, \eta(t))$ is the joint pdf of the random bivariate data $(t_1, \eta(t_1)) \ldots (t_J, \eta(t_J))$. Let $m(t)$ be an unknown regression function. Then the nonparametric regression model is

$$\eta(t_j) = m(t_j) + \epsilon_j; \quad j = 1, \ldots, J. \tag{3.3}$$

The errors (ϵ_j) satisfy $E(\epsilon_j) = 0, Var(\epsilon_j) = \sigma_\epsilon^2$ and $Cov(\epsilon_j, \epsilon_k) = 0$ for j $\neq$ k. The unknown regression function $m(t)$ will be derived as

$$m(t) = E[\eta(t)|T = t] = \int \eta(t) f\Big[\eta(t)|t\Big] d\eta(t) = \frac{\int \eta(t) f\Big[t, \eta(t)\Big] d\eta(t)}{\int f\Big[t, \eta(t)\Big] d\eta(t)},$$

where $m(t)$ is a ratio of two correlated random quantities. We use the product kernel density estimator technique to estimate the numerator and denominator seperately, i.e.,

$$\begin{aligned} \hat{f}\Big[t, \eta(t)\Big] &= \frac{1}{J h_t h_\eta} \sum_{j=1}^{J} K\Big(\frac{t - t_j}{h_t}\Big) K\Big(\frac{\eta(t) - \eta(t_j)}{h_\eta}\Big) \\ &= \frac{1}{J} \sum_{j=1}^{J} K_{h_t}\Big(t - t_j\Big) K_{h_\eta}\Big(\eta(t) - \eta(t_j)\Big). \end{aligned}$$

By using the symmetry of the kernel and transformation of variables, we have

$$\int \eta \hat{f}\Big[t,\eta(t)\Big]d\eta = \frac{1}{J}\int \eta \sum_{j=1}^{J} K_{h_t}\Big(t-t_j\Big)K_{h_\eta}\Big(\eta(t)-\eta(t_j)\Big)$$
$$= \frac{1}{J}\sum_{j=1}^{J} K_{h_t}\Big(t-t_j\Big)\eta(t_j).$$

For the denominator,

$$\int \hat{f}\Big[t,\eta(t)\Big]d\eta = \frac{1}{J}\sum_{j=1}^{J} K_{h_t}\Big(t-t_j\Big)\int K_{h_\eta}\Big(\eta(t)-\eta(t_j)\Big)d\eta$$
$$= \frac{1}{J}\sum_{j=1}^{J} K_{h_t}\Big(t-t_j\Big)$$
$$= \hat{f}(t).$$

Therefore, we have

$$\hat{m}(t) = \sum_{j=1}^{J} W_{h_t}\Big(t-t_j\Big)\eta(t_j), \tag{3.4}$$

where $W_{h_t}\Big(t-t_j\Big) = K_{h_t}\Big(t-t_j\Big)/\sum_{j=1}^{J} K_{h_t}\Big(t-t_j\Big)$ and $\sum_{j=1}^{J} W_{h_t}\Big(t-t_j\Big) = 1$. Estimator (3.4) is widely known as the Nadaraya-Watonson type kernel estimator. In smoothing estimation, the kernel works as a weighting function and bandwidth h as a smoothing parameter.

From equation (3.1), we have $h = \frac{D_1 D_2}{\hat{\eta}^q(t)}$, where $D_1 = q! e_{q+1,p+1}\Big[T_p^T(t)G(t;h)T_p(t)\Big]^{-1}$ and $D_2 = \sum_{j=1}^{J} T_{j,p}^T(t)(0,\ldots,K(\frac{t_j-t}{h}),\ldots,0)^T\widetilde{\eta}(t_j)$. From equation (3.4), we have $h = \frac{D_3}{D_4\hat{m}(t)}$, where $D_3 = \sum_{j=1}^{J} K\Big(\frac{t-t_j}{h}\Big)\eta(t_j)$, $D_4 = \sum_{j=1}^{J} K_{h_t}\Big(t-t_j\Big)$. Equating for h, which is known as the bandwidth and can be selected subjectively or by a data driven procedure, we have the following finite sample relationship between the two-step kernel smoothing estimator and the two-step local polynomial smoothing estimator

$$\hat{\eta}^q(t) = \frac{D_1 D_2 D_4}{D_3}\hat{m}(t). \tag{3.5}$$

3.3 Bandwidth Choices and Kernel Selection

In nonparametric regression, the bandwidth controls the smoothness and roughness of the smoothing estimator and the kernel works as the weighting function. The Performance of the kernel is measured by MISE (Mean Integrated Squared Error) or AMISE (Asymptotic MISE). The Epanechnikov kernel minimizes AMISE and is therefore considered optimal for comparison to other kernels. We use the Epanechnikov kernel in our application and simulation studies. Kernel selection is not as important as the choice of bandwidth.Two popular bandwidth selection techniques are "Leave-One-Subject-Out Cross Validation (LSCV)" and "Leave-One-Time-Point-Out Cross Validation (LTCV)." The LSCV procedure deletes observations one at a time while LTCV deletes observations at the time design points $\boldsymbol{t}=(t_1, \ldots, t_J)$. The bandwidths for (3.1) and (3.4) are selected by the LTCV procedure because our data in applications and simulations are binned to different time (age) points. The cross-validation criterion is

$$CV\left(h\right) = \sum_{j=1}^{J} \sum_{i \in \mathcal{S}_j} W_i \left\{ Y_i(t_j) - \widehat{\eta}^{(-j)}\left(t_j\right) \right\}^2, \tag{3.6}$$

where W_i be a weight function which could be either $1/(nm_i)$ or $1/N$ and $\widehat{\eta}^{(-j)}\left(t_j\right)$ is the nonparametric regression estimators of (3.1), (3.2) and (3.4) applied to the data at all time points except time point t_j. The CV choice of h is the one that minimizes $CV\left(h\right)$ over $h \geq 0$. Bandwidth choice plays a significant role in nonparametric regression. A subjective choice or wrong choice of very small ($h \to 0$) or very large ($h \to \infty$) bandwidth will produce undersmoothed or oversmoothed results. For very large choice of bandwidth, nonparametric estimates converge to the ordinary least squares fit of a straight line with severe bias problems. On the other hand, if bandwidth is very small, smoothing estimates will have large variances.

3.4 Bootstrap Confidence Band

A widely used inference approach for nonparametric analysis is the "resampling-subject" bootstrap suggested in Hoover *et al.* (1998). In the current context, we can obtain a pointwise bootstrap confidence interval for $\eta(t)$ by first obtaining B bootstrap samples through resampling the subjects of the cross sectional sample with replacement, and then computing B two-step smoothing estimators $\left\{\widehat{\eta}^b(t) : b = 1, \ldots B\right\}$ using (3.2) and (3.4) and each of the bootstrap samples. The lower and upper boundaries of the $[100 \times (1 - \alpha)]$th empirical quantile bootstrap pointwise confidence interval of $\widehat{\eta}(t)$

are the empirical lower and upper $[100 \times (\alpha/2)]$th percentiles based on the bootstrap estimators $\left\{\widehat{\eta}^b(t) : b = 1, \dots B\right\}$. Alternatively, if $SD\left\{\widehat{\eta}^b(t)\right\}$ is the empirical standard deviation of $\left\{\widehat{\eta}^b(t) : b = 1, \dots B\right\}$, the $[100 \times (1-\alpha)]$th normally approximated bootstrap pointwise confidence interval of $\widehat{\eta}(t)$ is

$$\widehat{\eta}(t) \pm Z_{1-\alpha/2} \times SD\left\{\widehat{\eta}^b(t)\right\},$$

where $Z_{1-\alpha/2}$ is the $[100 \times (1-\alpha/2)]$th percentile of the standard normal distribution.

4 Application to Bangladesh Temperature Data

We apply our methods to temperature data in degrees Celsius from the capital city of Bangladesh. These data were recorded on each day by the Bangladesh Meteorological Department from 1990 to 2015. From this data, we computed the minimum, maximum and average temperature for each month and then further computed the minimum, maximum and average temperature of the year. We have $J = 26$ distinct time design points $\{t_1, t_2, \dots, t_{26}\} = \{1990, 1991, \dots, 2015\}$. Thus, for a given $1 \le j \le J = 26$, we denote $Y_{Mean}(t_j)$, $Y_{Max}(t_j)$ and $Y_{Min}(t_j)$ as the mean, maximum and minimum temperature at year t_j. Applying the two-step local linear estimator of (3.2) and kernel estimator (3.4) to the observed data $\left\{Y_{Mean}(t_j), t_j; 1 \le j \le J, 1 \le i \le n\right\}$, $\left\{Y_{Max}(t_j), t_j; 1 \le j \le J, 1 \le i \le n\right\}$ and $\left\{Y_{Min}(t_j), t_j; 1 \le j \le J, 1 \le i \le n\right\}$, we compute the smoothing estimators of minimum, maximum and average temperature on the entire time design points $\{t_1, t_2, \dots, t_{26}\} = \{1990, 1991, \dots, 2015\}$.

Figure 1 shows the local linear smoothing estimates and kernel smoothing estimates of $Y_{Mean}(t_j)$, $Y_{Max}(t_j)$ and $Y_{Min}(t_j)$ and their corresponding bootstrap pointwise 95% confidence interval based on B=1000 bootstrap replications. A bandwidth of $h = 2.85$ and the Epanechnikov kernel were used. This bandwidth was chosen by minimizing the LTCV scores. Bootstrap confidence bands have been constructed to demonstrate that bandwidth choice is made correctly and also to see which smoothing estimator has narrower confidence band. An incorrect choice of bandwidth would produce smoothing estimates that would fall out of the bootstrap confidence bands. In Figure 1, dots represent the raw estimates, solid black lines represent smoothing estimates and dotted lines represent the 95% pointwise bootstrap confidence bands. By looking at the figures, we can see that the kernel smoothing estimator is a little rough compare to the local linear smoothing estimator and local linear smoothing estimator produces narrower bootstrap confidence bands. In part (e) and (f) of Figure 1, we see that average temperature is slowly increasing with time, clearly indicating the evidence that global warming is a reality. A close look at Figure 1 (part (e) and part (f)) and Table 1 tells that there is a 0.45^o Celsius increase in average temperature in a 26 year period from 1990 to 2015. Table 1 shows nonparametric raw estimators (Max, Min and Mean), two-step kernel smoothing estimators (Max.k, Min.K and Mean.K) and two-step local linear smoothing estimators (Max.L, Min.L and Mean.L) for $Y_{Max}(t_j)$, $Y_{Min}(t_j)$ and $Y_{Mean}(t_j)$ between 1990 and 2015. In Table 1, Max.K stands for two-step kernel smoothing estimates of $Y_{Max}(t_j)$, Max stands for $Y_{Max}(t_j)$ and Max.L stands for two-step local linear smoothing estimates of $Y_{Max}(t_j)$. Similar abbreviations stand for the minimum and average temperatures. Tabular representation of pointwise bootstrap confidence band have been omitted to avoid redundancy.

5 Simulation

In this section, we conduct a set of simulation studies to assess the performance of the two two-step smoothing estimators. We compare the performance of these two estimators by computing the ratio of MSE from the local linear smoothing estimator to that of the kernel smoothing estimator. Two cross-sectional and two longitudinal frameworks of data simulation are considered. Cross-sectional designs have time-variant fixed parameters from parametric family and time-variant variable parameters from a parametric family. For longitudinal simulation, we consider consider both balanced and unbalanced longitudinal designs. For simulating data from a cross-sectional design, we generate a sample of $m = 1000$ subjects at 25 different time points. For the longitudinal design, we considered 100 time points because of repeated measurements. We then repeat this procedure for $S = 500$ times to generate a set of 500 random samples. Within each simulated sample, we first compute the raw estimates of the time-variant parameters and then in the second step, we compute the smoothing estimates using local linear smoothing estimators and kernel smoothing estimators. The Epanechnikov kernel and a bandwidth of $h = 2.85$ is used for all simulation designs. We finally compute the MSE and ratio of MSE of these two estimators. Results from Table 2 to Table 5 show that the local linear smoothing estimator outperform the kernel smoothing estimator by producing smaller MSE in all time points except at a few boundary points.

Simulation − 1 : For generating data from the time-variant fixed parametric model, we consider three discrete probability models and three continuous probability models with fixed parameter values at each time point. Probability models together with their fixed parameter values are shown on Table 2.

Simulation − 2 : The same probability models are used as in simulation design 1,

Table 1: Raw estimate, local linear smoothing estimate and Kernel smoothing estimate of maximum, minimum and average temperature from the capital city of Bangladesh from 1990 to 2015. The Epanechnikov kernel and the bandwidth $h = 2.85$ are used for the smoothing estimators.

Year	Max.K	Max	Max.L	Min.K	Min	Min.L	Mean.K	Mean	Mean.L
1990	31.04	31.2	31.02	16.71	15.7	16.89	25.75	25.75	25.7
1991	30.99	31.1	31.01	16.78	18.5	16.83	25.73	25.8	25.71
1992	30.95	30.9	30.99	16.85	15.7	16.75	25.71	25.83	25.72
1993	30.93	30.5	30.98	16.89	15.7	16.66	25.7	25.39	25.73
1994	30.94	31.2	30.98	16.82	20.1	16.57	25.72	25.7	25.75
1995	30.95	31.2	30.98	16.58	15.7	16.48	25.74	25.92	25.77
1996	30.98	30.5	30.97	16.25	15.7	16.4	25.76	25.89	25.79
1997	31.01	31.1	30.97	15.97	15.7	16.32	25.8	25.36	25.81
1998	31.03	31.2	30.97	15.85	15.7	16.26	25.84	25.89	25.84
1999	31.01	31.2	30.97	15.89	15.7	16.2	25.88	26.29	25.86
2000	30.96	31.2	30.96	16.04	15.7	16.15	25.9	25.74	25.89
2001	30.91	30.1	30.96	16.22	15.7	16.12	25.9	25.95	25.92
2002	30.92	31.2	30.95	16.3	18.5	16.09	25.9	25.83	25.95
2003	30.96	30.9	30.94	16.22	15.7	16.07	25.93	25.76	25.98
2004	31	31.2	30.93	16.04	15.7	16.06	25.99	25.89	26
2005	31.01	30.9	30.92	15.86	15.7	16.06	26.06	26.25	26.03
2006	30.97	31.2	30.9	15.76	15.7	16.08	26.1	26.48	26.06
2007	30.9	31.2	30.88	15.73	15.7	16.12	26.12	25.72	26.09
2008	30.84	30.1	30.86	15.77	15.7	16.18	26.16	25.89	26.11
2009	30.83	30.9	30.84	15.88	15.7	16.27	26.19	26.54	26.13
2010	30.84	31.2	30.81	16.1	15.7	16.39	26.2	26.58	26.15
2011	30.84	30.5	30.77	16.39	15.7	16.55	26.18	25.84	26.16
2012	30.81	31.2	30.73	16.67	18.5	16.75	26.15	26.1	26.17
2013	30.76	30.9	30.68	16.84	15.7	16.99	26.14	26.1	26.18
2014	30.71	30.1	30.62	16.88	18.5	17.28	26.14	26.22	26.18
2015	30.67	30.9	30.55	16.83	15.7	17.6	26.15	26.16	26.18

but the values of the parameters are changed randomly within a pre-specified interval. Table 3 specifies the varying parameter values from the time-variant parametric models.

Simulation – 3 : For simulating data under the balanced longitudinal framework, we use “Orthodont" data set available in the R package “nlme”. The Number of measurements for each subject is four and these measurements are equidistant along time. We use a linear mixed effect model to simulate data where Distance is considered as response variable and Age and Sex are used as predictors. Simulated data is split by age and then used to compute average distance, which is the first step in the two-step estimation procedure. In the second step, we apply smoothing estimators. Table 4 gives the ratio of MSEs in all 100 time points.

Simulation – 4 : We simulate data from an unbalanced longitudinal framework with the response variable having the AR(1) correlation structure. For this design, the number of subjects could be 2 to 10 and their measurements are not equidistant. To simulate AR(1) errors, we use the *arima.sim* function from R package “MASS" with $\rho = .4$ and $\sigma = 1.5$. We denote this error as e_{ij}. We then generate the response variable y_{ij} from the model

$$y_{ij} = \beta_0 + u + (\beta_1 + v)\log(i) + e_{ij} + age,$$

where u and v are the first and second columns generated from a bivariate normal distribution with $\mu = (0, 0)$ and covariance matrix ($\sigma_1^2 = 6.25, \sigma_2^2 = 4, \sigma_{12} = 1.5$). i is the observation number and age is randomly generated from uniform distribution on [0,10]. $\beta_0 = 1, \beta_1 = 6$. After generating the data, we split the data by age and computed the mean of the response variable at each time point. We then applied smoothing estimators to compute the smoothing estimates of mean. The ratios of the MSEs are given in the table 5.

6 Discussion

We proposed a class of time-varying parametric models and time-variant regression models for smoothing estimation of the parameters by two nonparametric smoothing estimators. The mathematical relationship between the two smoothing estimator has been established. Parameters from time-variant semi-parametric regression models can also be smoothed by these smoothing estimators. Odds Ratios from logistic models, hazard ratios from Cox models and competing risk models and C-statistics from any regressionrnel smoothing estimator. models can also be smoothed by these smoothing techniques when time-variant data are available. Application and simulation results show that the local linear smoothing estimator outperforms the ke There

Table 2: Ratio of MSE of the local linear smoothing estimates to the kernel smoothing estimates from time-varying fixed parametric models (simulation design-1).

Time Points	Geometric p(t)=0.30	Binomial p(t)=0.6 n=10	Poisson $\lambda(t) = 4$	Exponential $\lambda(t) = 1/7$	Normal $\mu(t) = 10$ $\sigma = 2$	Cauchy $\mu(t) = 17$ $\theta = 2$
1	1.1319	1.1206	1.1334	1.1566	1.1387	1.1187
2	1.0369	1.035	1.0389	1.0587	1.027	1.0207
3	0.9188	0.9142	0.9174	0.9243	0.8929	0.8782
4	0.7892	0.7614	0.7818	0.7696	0.7553	0.7316
5	0.6906	0.6278	0.6619	0.6395	0.6434	0.634
6	0.6305	0.5356	0.5566	0.5447	0.5642	0.5745
7	0.5806	0.4761	0.4734	0.4727	0.5044	0.5222
8	0.5279	0.4398	0.4274	0.423	0.4507	0.4645
9	0.4863	0.4193	0.4106	0.3981	0.4093	0.4155
10	0.4648	0.4032	0.4021	0.3868	0.3954	0.3956
11	0.4579	0.3835	0.3899	0.379	0.4069	0.4029
12	0.4547	0.3669	0.3744	0.3709	0.4156	0.4109
13	0.4521	0.3633	0.3613	0.3628	0.3983	0.4014
14	0.4587	0.3732	0.3624	0.3634	0.3724	0.3874
15	0.4756	0.3939	0.3812	0.3798	0.3627	0.3872
16	0.4947	0.4177	0.4078	0.4081	0.3768	0.4102
17	0.5194	0.4418	0.4305	0.4391	0.4194	0.4551
18	0.5637	0.4815	0.4555	0.4746	0.4965	0.5125
19	0.6363	0.55	0.5115	0.5318	0.5967	0.5706
20	0.7343	0.6536	0.6271	0.6336	0.7051	0.6402
21	0.8624	0.8027	0.8057	0.7923	0.8364	0.7704
22	1.0338	0.9937	1.011	1.0012	1.0059	0.9964
23	1.2325	1.2108	1.1971	1.2248	1.2023	1.2471
24	1.4159	1.4412	1.3488	1.393	1.4025	1.413
25	1.5735	1.6689	1.4848	1.4949	1.5935	1.5179

Table 3: Ratio of MSE of the local linear smoothing estimates to the kernel smoothing estimates from time-varying variable parametric models (simulation design-2).

Time points	Geometric $p(t) \in [.2, .8]$	Binomial $p(t) \in [.4, .7]$	Poisson $\lambda(t) \in [4, 10]$	Exponential $\lambda(t) \in [.3, .7]$	Normal $\mu(t) \in [10, 20]$ $\sigma = 2$	Cauchy $\mu(t) \in [7, 27]$ $\theta = 2$
1	1.0273	1.0712	1.032	1.0035	1.0112	1.1031
2	0.9994	1.0449	1.0102	1.001	1.0051	1.0095
3	0.9843	1.0062	0.9897	0.9986	0.9971	0.9193
4	0.9791	0.9429	0.9721	0.9972	0.986	0.7954
5	0.975	0.8718	0.9611	0.9974	0.9749	0.6535
6	0.9655	0.8255	0.9578	0.9986	0.9705	0.5521
7	0.9566	0.8192	0.9589	0.9998	0.9753	0.5011
8	0.9533	0.8388	0.9632	0.9999	0.9836	0.4757
9	0.9523	0.8438	0.9733	1.0000	0.9861	0.4585
10	0.9502	0.8095	0.9911	0.9999	0.9785	0.4403
11	0.9423	0.7689	0.9997	0.9998	0.9644	0.4071
12	0.9284	0.7551	0.9998	0.9997	0.9515	0.3657
13	0.9169	0.7684	0.9999	0.9997	0.9476	0.3484
14	0.9155	0.7943	1.0000	0.9997	0.9571	0.3671
15	0.9257	0.8111	0.996	0.9989	0.9755	0.4022
16	0.9441	0.8037	0.9898	0.9983	0.9907	0.4331
17	0.9603	0.7834	0.9819	0.9981	0.9949	0.4692
18	0.9632	0.7727	0.9691	0.9982	0.9916	0.5187
19	0.9567	0.7827	0.9576	0.9984	0.9875	0.5788
20	0.9553	0.8182	0.9572	0.999	0.985	0.6586
21	0.9681	0.8865	0.971	1.0000	0.9843	0.7892
22	0.9976	0.9771	0.9942	1.0011	0.9865	0.9834
23	1.042	1.0477	1.0212	1.0019	0.9929	1.197
24	1.0939	1.0674	1.0498	1.0025	1.0044	1.3952
25	1.1481	1.0618	1.0823	1.0036	1.022	1.5921

Table 4: Ratio of MSE of the local linear smoothing estimates to the kernel smoothing estimates from balanced longitudinal design (simulation design-3).

Time Points	MSE Ratio	Time Points	MSE Ratio	Time Points	MSE Ratio	Time Points	MSE Ratio
0.1	1.158	2.6	0.4456	5.1	0.5183	7.6	0.5499
0.2	1.0832	2.7	0.4619	5.2	0.5459	7.7	0.5305
0.3	0.9604	2.8	0.4916	5.3	0.587	7.8	0.5001
0.4	0.8249	2.9	0.5168	5.4	0.6149	7.9	0.4949
0.5	0.7065	3.0	0.5193	5.5	0.609	8.0	0.5187
0.6	0.595	3.1	0.5142	5.6	0.5879	8.1	0.5571
0.7	0.5115	3.2	0.5172	5.7	0.5576	8.2	0.5868
0.8	0.4688	3.3	0.5168	5.8	0.5156	8.3	0.5873
0.9	0.4537	3.4	0.5105	5.9	0.4826	8.4	0.5732
1.0	0.4596	3.5	0.5186	6.0	0.4795	8.5	0.5696
1.1	0.4749	3.6	0.5477	6.1	0.5076	8.6	0.5732
1.2	0.4687	3.7	0.5753	6.2	0.5466	8.7	0.5734
1.3	0.4389	3.8	0.5779	6.3	0.5697	8.8	0.5682
1.4	0.4156	3.9	0.5634	6.4	0.575	8.9	0.5543
1.5	0.4143	4.0	0.5563	6.5	0.5703	9.0	0.5351
1.6	0.4355	4.1	0.5571	6.6	0.559	9.1	0.5322
1.7	0.4653	4.2	0.5436	6.7	0.555	9.2	0.5691
1.8	0.4792	4.3	0.516	6.8	0.561	9.3	0.6437
1.9	0.4774	4.4	0.4975	6.9	0.567	9.4	0.723
2.0	0.4745	4.5	0.5024	7.0	0.5781	9.5	0.7756
2.1	0.4702	4.6	0.5295	7.1	0.5942	9.6	0.8055
2.2	0.4649	4.7	0.5601	7.2	0.5893	9.7	0.8674
2.3	0.4601	4.8	0.5642	7.3	0.5553	9.8	1.0105
2.4	0.4524	4.9	0.5388	7.4	0.5297	9.9	1.1718
2.5	0.4448	5.0	0.5156	7.5	0.5371	10	1.2097

Table 5: Ratio of MSE of the local linear smoothing estimates to the kernel smoothing estimates from unbalanced longitudinal design (simulation design-4).

Time Points	MSE Ratio	Time Points	MSE Ratio	Time Points	MSE Ratio	Time Points	MSE Ratio
0.1	1.0639	2.6	0.4449	5.1	0.5251	7.6	0.475
0.2	1.1071	2.7	0.4433	5.2	0.4814	7.7	0.4796
0.3	1.0747	2.8	0.4196	5.3	0.4418	7.8	0.473
0.4	0.9169	2.9	0.388	5.4	0.4086	7.9	0.4731
0.5	0.7551	3.0	0.3712	5.5	0.3765	8.0	0.4808
0.6	0.6571	3.1	0.3737	5.6	0.3718	8.1	0.4864
0.7	0.6056	3.2	0.3872	5.7	0.4069	8.2	0.4846
0.8	0.564	3.3	0.4031	5.8	0.4528	8.3	0.4743
0.9	0.5085	3.4	0.4066	5.9	0.4647	8.4	0.4559
1.0	0.4619	3.5	0.3966	6.0	0.4452	8.5	0.4378
1.1	0.4484	3.6	0.3918	6.1	0.4283	8.6	0.4271
1.2	0.4389	3.7	0.3985	6.2	0.4292	8.7	0.4219
1.3	0.4011	3.8	0.4158	6.3	0.4473	8.8	0.4314
1.4	0.362	3.9	0.4327	6.4	0.4807	8.9	0.4609
1.5	0.3523	4.0	0.4305	6.5	0.5105	9.0	0.4752
1.6	0.3697	4.1	0.4112	6.6	0.4993	9.1	0.4527
1.7	0.3961	4.2	0.4022	6.7	0.4567	9.2	0.4322
1.8	0.4155	4.3	0.4185	6.8	0.4226	9.3	0.4483
1.9	0.4226	4.4	0.4403	6.9	0.411	9.4	0.5095
2.0	0.4221	4.5	0.4408	7.0	0.4187	9.5	0.6083
2.1	0.4177	4.6	0.4354	7.1	0.4409	9.6	0.724
2.2	0.4125	4.7	0.4483	7.2	0.4691	9.7	0.8508
2.3	0.4148	4.8	0.4732	7.3	0.4795	9.8	1.022
2.4	0.4247	4.9	0.5013	7.4	0.4667	9.9	1.2096
2.5	0.4354	5.0	0.53	7.5	0.4627	10.0	1.2788

are a number of theoretical and methodological aspects that warrant further investigation. Theoretical studies are warranted to investigate the properties of other smoothing methods, such as the global smoothing methods through splines, wavelets and other basis approximations, and their corresponding asymptotic inference procedures.

References

Chowdhury, M., Wu, C. and Modarres, R. (2016). Local Box-Cox transformation on time Varying parametric model for smoothing estimation of conditional CDF with longitudinal data. *Journal of Statistical Computation and Simulation (To appear)*.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of American Statistical Association*, **74**, 829-836.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of American Statistical Association*, **87**, 998-1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiency. *Annals of Statistics*, **21**, 196-216.

Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Annals of Statistics*, **20**, 2008-2036.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*, Chapman and Hall: London.

Fan, J. and Zhang, J.T.(2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society, Ser. B*, **62**, 303-322.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, **85**, 809-822.

Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability & Its Applications*, **9**, 141-142.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics*, **22(3)**, 1346-1370.

Stone, C.J. (1977). Consistent nonparametric regression. *Annals of Statistics*, **5**, 595-645.

Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators, *Annals of Statistics*, **8**, 1348-1360.

Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics*, **10**, 1040-1053.

Wu, C. O., Tian, X. and Yu, J. (2010). Nonparametric estimation for time-varying transformation models with longitudinal data, *Journal of Nonparametric Statistics*, **22**, 133-147.

Wu, C. O. and Tian, X. (2013a). Nonparametric estimation of conditional distribution functions and rank-tracking probabilities with longitudinal data. *Journal of Statistical Theory and Practice*, **7** 1-26.

Wu, C. O. and Tian, X. (2013b). Nonparametric estimation of conditional distribution functions and rank-tracking probabilities with time-varying transformation models in longitudinal studies. *Journal of the American Statistical Association*, **108**, No. 503.

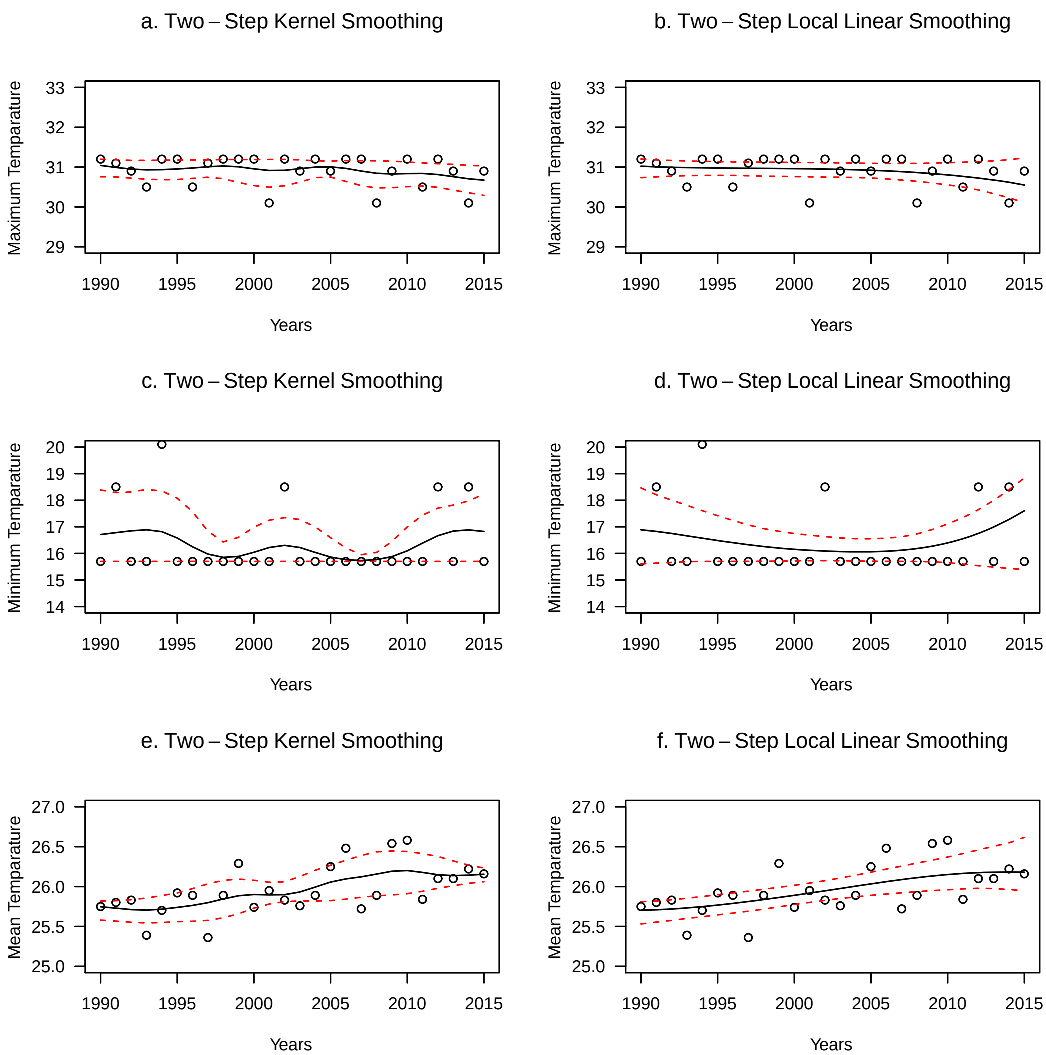

Figure 1: Raw estimators (dots), smoothing estimators (solid curves), and bootstrap pointwise 95% confidence intervals (dashed curves, B=1000 bootstrap replications) for minimum, maximum and average temperature.