NONPARAMETRIC ESTIMATION OF TIME-VARIANT PARAMETRIC MODELS WITH APPLICATION TO CROSS-SECTIONAL DATA

Mohammed Chowdhury*

In this article, two estimation approaches based on age-specific parametric model have been proposed and a comparative study between them has been studied. We assume that outcome variable follows a parametric model, but the parameters are smooth function of time (age). Our estimation is based on a two-step smoothing method, in which we first obtain the raw estimators of the parameters at a set of disjoint time points, and then compute the final estimators at any time by smoothing the raw estimators. We derived asymptotic properties such as asymptotic biases, variances and mean squared error (MSE) for the local polynomial smoothed estimator and kernel smoothing estimator for the parameter of the time-variant parametric model. A mathematical relationship is established between two asymptotic MSEs. Mathematical relationship between two smoothing estimators has also been established. Applications of our two-step estimation method have been demonstrated through a large demographic study to estimate fecundability. Theoretical results on coverage of bootstrap confidence intervals for these smoothing estimators have been derived. Finite sample properties of our procedures are investigated by a simulation study.

Key words and phrases: Cross validation and bandwidth, fecundability, geometric model, kernel smoothing, local Polynomials, two-step smoothing.

1. Introduction

Cross sectional data mainly arise from retrospective survey. Exiting statistical models involving cross sectional data mainly focus on regression models based on conditional mean and conditional variance-covariance structure. These methods, although popular in practice, may not be appropriate for modeling, estimating and predicting parameter value on entire time design points. In order to model and estimate parameters directly on entire time design points, we have adopted a two-step smoothing method. Motivated by the works of Hoover et al. (1998) on time-varying coefficient models, we propose in this paper a structural nonparametric approach for the estimation of parameters, and show that, when 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 propose a two-step smoothing procedure which first obtains the raw estimators of the parameters based on the time-varying parametric family at a set of distinct time

*Department of Statistics and Analytical Sciences, Kennesaw State University, Georgia, U.S.A. Email: chowdhury@kennesaw.edu
points, and then computes the final estimators at any time point by smoothing the available raw estimators 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 be implemented in practice. For the practical properties, we demonstrate the implications of our structural nonparametric approach through an application to the Bangladesh Demographic and Health Survey (BDHS, 2011) data BDHS (2011), and investigate the finite sample properties of our procedures through a simulation study. For the theoretical properties, we derive the asymptotic distributions of the raw estimators and the asymptotic expressions for the biases, variances and mean squared errors for the two-step local polynomial estimators and kernel smoothing estimators. We have also established mathematical relationship between two asymptotic MSEs and two smoothing estimators.

The problem of modeling and estimating parameter or it functionals is well motivated by BDHS-2011, which has one of the main objectives of evaluating the fertility performance of woman. Fertility rate, fecundity and fecundability are three important interrelated parameters. Fecundity means biological capacity of women to conceive and the women who have this capability is known as fecund. Fecundability means monthly chance of conception. If the couples do not use any contraceptive methods before their first birth, then fecundability is known as mean natural fecundability. In this article, smoothing estimation of fecundability has been developed from a parametric family to demonstrate the application of our procedure.

Fecundability has an inverse relation to the time interval required to conceive (Sheps and Menken (1973)). This time interval is known as the Conception Wait (CW). CW is computed as, CW = date of first birth – 9 months – date of marriage. CW and fecundability are two interrelated fertility parameters and are also regarded as the most direct measures of fertility performance of a population. According to the CDC (Centers for Disease Control and Prevention), infertility is defined as not being able to get pregnant after one year of unprotected intercourse. In USA, About 6% of married women 15–44 years of age are unable to get pregnant after one year of unprotected intercourse. Also, about 12% of women 15–44 years of age in the United States have difficulty getting pregnant or carrying a pregnancy to term, regardless of marital status.

Fecundability and marital fertility are linked through the following chain of variables: frequency of unprotected coitus $\rightarrow$ fecundability $\rightarrow$ exposure interval $\rightarrow$ birth interval $\rightarrow$ marital fertility rate (Bongaarts (1975)). Fecundability is considered as the transition probability for the passage from the susceptible state to pregnancy (Perrin and Sheps (1964)). In a homogeneous population, fecundability is equal to the reciprocal of its mean conception delay (Sheps and Menken (1973)) but for heterogeneous populations, the mean fecundabilities are usually modeled on two parameters (Porter and Parker (1964), Jain (1969), and Majumdar and Sheps (1970)). Chowdhury and Umbach (2012) also used three Bayesian methods to estimate fecundability for the heterogeneous
populations. Pandey and Yadav (2013) developed probability model of waiting time to first conception under Bayesian framework. These methods are not adequate for estimating and predicting fecundabilities on entire time design points. Alternatively, none attempted to develop smoothing estimators for predicting fecundabilities in the literature of population studies and demography.

In the main results, we describe the data structure, the conditional probability model for estimating fecundability and our time-varying parametric models in Section 2, and present our estimation methods in Section 3. For the numerical results, we apply our estimation and inference procedures to the BDHS data in Section 4 and present our simulation results in Section 5. We develop the asymptotic properties of our estimators in Section 6. Finally, we briefly discuss in Section 7 some further implications and extensions of the theory and applications for the estimation of the fecundabilities. Further details of the theoretical derivations are given in the Appendix.

2. Data and models
2.1. Data structure

This study utilizes data from the Bangladesh Demographic and Health Survey (BDHS, 2011), which is a part of the worldwide Demographic and Health Survey Program. ICF International of Calverton, Maryland, USA provided technical assistant to the project as part of its international Demographic and Health Surveys program (MEASURE, DHS), and financial assistance was provided by the U.S. Agency for International Development (USAID). A detailed description of the methodology of the data collection including sample design for the survey can be found in BDHS (2011). DHS (Demographic and Health Survey) data from other countries can also be found at http://dhsprogram.com/data/available-datasets.cfm.

For smoothing estimation of natural fecundability, we have extracted 7780 women out of 17842 women who did not use any contraception before giving first birth. Table 1 gives the conception wait and the number of cumulative conceptions within six years of their marriage. There are two rounding approaches for conception wait. First, any conception that takes place between day 0th and day 15th are considered as zero month, any conception that takes place between day 16th and day 45th are considered as one month and so on until 72th month. Second, conceptions that takes places between day 0th and 30th are considered as first month, any conception that takes place between any days of second calendar month is considered as second month and so on. We have considered the second rounding approach to incorporate geometric probability model (2.2) to estimate fecundability. The estimates of fecundabilities from the conception wait obtained by two rounding approaches differ very little. The main reasons of using BDHS, 2011 for estimating fecundabilities are: number of couples who never used contraception before their first birth, low chances of having intercourse before marriage and availability of date of marriage. These three characteristics of the couples give fair estimates of mean natural fecundabilities. In our analysis,
Table 1. Conception wait and the number of conceptions.

| Conception wait in Months | Number of Conceptions |
|---------------------------|-----------------------|
| 1-12.0                    | 3367                  |
| 1-24.0                    | 5335                  |
| 1-36.0                    | 6556                  |
| 1-48.0                    | 7219                  |
| 1-60.0                    | 7593                  |
| 1-72.0                    | 7780                  |

we consider minimum age at first conception as 14 years and consequently first birth as 14 years and 9 months. Any women, who conceived earlier than 14 years of age are ignored from the analysis to avoid memory bias.

Within the BDHS sample, we have \( n \) independent subjects and for each subject we have only one observation at one of the \( J \) disjoint “time design points” \( t = (t_1, \ldots, t_J)^T \). For example, \( S_j \) is the set of \( n_j \) woman who conceived at age \( t_j \), and \( \{Y_i(t_j), i = 1, \ldots, n_j; j = 1, \ldots, J\} \) are their conception wait time. In population and biomedical studies, \( t \) is often obtained by rounding off age or other time variables within an acceptable accuracy range. So, our cross sectional sample could be written as \( Z = \{Y_i(t_j), t_j, i \in S_j\} \) where \( Y_i(t_j) \) is the random month of conception wait for the \( i \)th women at \( j \)th time point \( (j = 1, \ldots, J; i = 1, \ldots, n_j) \). \( t_j \in \tau \), where \( \tau \) is the time interval containing the time range of interest and \( S_j \) is the set of subjects which have observations at time point \( t_j \). The time points in \( t \) are specified by rounding up the conception ages of the BDHS subjects.

2.2. Time-varying parametric models

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

\[
F_{\theta(t)} = \{F_{t,\theta(t)}(\cdot); \theta(t) \in \Theta\},
\]

where \( \theta(t) \) is the vector of time-varying parameters which belong to an open Euclidean space \( \Theta \). For special case of time-varying geometric model for CW to estimate fecundability at time points \( t = (t_1, \ldots, t_J) \), we have

\[
P\{Y(t_j) = y_i(t_j) \mid t_j, \theta(t_j)\} = \theta(t_j)\{1 - \theta(t_j)\}^{y_i(t_j) - 1}.
\]

Where \( \theta(t_j) \) is the monthly probability of conception (Fecundability) at \( t_j \), \( y_i(t_j) \) is the CW of the \( i \)th woman at first conception age \( t_j \). Chowdhury et al. (2017a, b) showed that when time-varying parametric assumption holds, a smoothing estimator of conditional distribution function, which utilizes the local parameter structures, could be a superior smoothing estimators under log transformation or local Box-Cox transformation.

The conception data seems iid for each couple. But they are not biologically iid. For example, women who are in their early age of menstruation (age 13–age
or at the age of menopause (age 45 or above) are less fertile than other women. Also frequency of coitus and timing of ovulation (which last up to 72 hours for each menstrual cycle) vary for each couple. For this, age specific homogeneous cohort are made for estimating fecundability at first step. In the second step, smoothing estimators are applied on them.

3. Two-step estimation

We develop here two two-step smoothing methods for the estimation of the parameter (fecundabilities) by first computing the raw estimates of $\theta(t_j)$ for all $j = 1, \ldots, J$, and then derive the smoothing estimates of $\theta(t)$ for any $t \in \tau$ by applying smoothing procedure over the raw estimates. The two-step smoothing approach is computationally simple and does not need correlation assumptions across different time points.

3.1. Raw estimates

First, we derive the raw estimates $\tilde{\theta}(t_j)$ of $\theta(t_j)$ using observations at time $t_j \in t$. For doing this, we suppose that we have enough observations $n_j$ at $t_j$, so that $\theta(t_j)$, $t_j \in t$, can be estimated by the maximum likelihood estimators (MLE) $\tilde{\theta}(t_j)$ using the subjects in $S_j$.

In practice, these raw estimators require the number of observations $n_j$ at $t_j$ to be sufficiently large, so that 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 round off or binning approach has been used by Fan and Zhang (2000), and Chowdhury et al. (2017a, b).

3.2. Smoothing estimators of parameter

3.2.1. Rationales of smoothing

There are two reasons to use the smoothing step in addition to the raw estimates. First, the raw estimates are only for the coefficients or fecundabilities at time points in $t$, while the smoothing step yields curve estimates over the entire time range $\tau$. Second, the raw estimates usually have excessive variations, so that their values may change dramatically among adjacent time points in $t$ due to the size and values of the samples. Given that spiky estimates may not have meaningful interpretations, the smoothing step should be used to reduce the variation by sharing information from the adjacent time points. In this section, we will develop two smoothing estimators and in Sections 5 and 6, we will compare them and establish two mathematical relationship.

3.2.2. Local polynomial smoothing estimators

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

$$\theta(t) \approx \sum_{q=0}^{p} \beta_q(a_0)(t - a_0)^q$$
for \( t \) in some neighborhood of \( a_0 \). We can treat the raw estimates \( \tilde{\theta}(t_j) \) as the “observations” of \( \theta(t_j) \) at \( t_j, j = 1, \ldots, J \), and obtain the \( p \)th local polynomial estimators by minimizing

\[
\sum_{j=1}^{J} \left\{ \tilde{\theta}(t_j) - \sum_{q=0}^{p} \beta_q(t)(t_j - t)^q \right\}^2 K_h(t_j - t),
\]

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 \( \tilde{\theta}(t) = (\tilde{\theta}(t_1), \ldots, \tilde{\theta}(t_J))^T, \beta(t) = (\beta_0(t), \ldots, \beta_p(t))^T, G(t; h) = \text{diag}\{K_h(t_j - t)\} \) with \( j \)th 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 \( j \)th row given by \( T_{j,p}(t) = (1, t_j - t, \ldots, (t_j - t)^p) \). The local polynomial estimators \( \hat{\beta}_q(t) \) minimize

\[
Q_C[\beta(t)] = [\tilde{\theta}(t) - T_p(t)\beta(t)]^TG(t; h)[\tilde{\theta}(t) - T_p(t)\beta(t)].
\]

The \( p \)th order local polynomial estimator of \( \theta^{(q)}(t) \) based on \( \tilde{\theta}(t_j) \), which minimizes \( Q_C[\beta(t)] \), is

\[
(3.1) \quad \hat{\theta}^{(q)}(t) = \sum_{j=1}^{J} \{W_{q,p+1}(t_j, t; h)\tilde{\theta}(t_j)\}
\]

where \( W_{q,p+1}(t_j, t; h) = q!e_{q+1,p+1}[T_p^T(t)G(t; h)T_p(t)]^{-1}[T_p^T(t)G_j(t; h)] \) 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 \( \beta(t) \), we have \( \hat{\beta}(t) = (\hat{\beta}_0(t), \ldots, \hat{\beta}_p(t))^T \) and \( \hat{\theta}^{(q)}(t) = \hat{\beta}_q(t)q! \) is an estimator for \( \theta^{(q)}(t) \), \( q = 0, 1, \ldots, p \). For local polynomial fitting \( p-q \) should be taken to be odd as shown in Ruppert and Wand (1994) and Fan and Gijbels (1996). When \( p = 1 \), we get the local linear estimator \( \hat{\theta}_L(t) = \hat{\beta}_0(t) \) of \( \theta(t) \) based on (3.1) and the equivalent kernel function \( W_{0,2}(t_j, t; h) \). So, the local linear estimator is

\[
(3.2) \quad \hat{\theta}_L(t) = \hat{\theta}^{(0)}(t \mid x).
\]

### 3.2.3. Kernel smoothing estimators

Suppose that we have random sample of bivariate data \((t_1, \theta(t_1)) \ldots (t_J, \theta(t_J))\) from a joint pdf \( f(t, \theta(t)) \). Let \( m(t) \) be an unknown regression function. Then the nonparametric regression model is

\[
(3.3) \quad \theta(t_j) = m(t_j) + \epsilon_j; \quad j = 1, \ldots, J.
\]

The errors \( (\epsilon_j) \) satisfy \( E(\epsilon_j) = 0 \), \( \text{Var}(\epsilon_j) = \sigma^2_\epsilon \) and \( \text{Cov}(\epsilon_j, \epsilon_k) = 0 \) for \( j \neq k \). The unknown regression function \( m(t) \) will be derived as follows:

\[
m(t) = E[\theta(t) \mid T = t] = \int \theta(t)f[\theta(t) \mid t]d\theta(t) = \frac{\int \theta(t)f[t, \theta(t)]d\theta(t)}{\int f[t, \theta(t)]d\theta(t)}
\]
$m(t)$ is a ratio of two correlated random quantities. We use product kernel density estimator technique to estimate the numerator and denominator separately. i.e;

$$
\hat{f}(t, \theta(t)) = \frac{1}{Jh_t h_\theta} \sum_{j=1}^{J} K\left(\frac{t - t_j}{h_t}\right) K\left(\frac{\theta(t) - \theta(t_j)}{h_\theta}\right)
= \frac{1}{J} \sum_{j=1}^{J} K_{h_t}(t - t_j)K_{h_\theta}(\theta(t) - \theta(t_j)).
$$

By using the property of symmetric kernel and transformation of variable, we have

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

For the denominator, we have

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

Therefore, we have

$$
(3.4) \quad \hat{m}(t) = \sum_{j=1}^{J} W_{h_t}(t - t_j) \tilde{\theta}(t_j).
$$

Where $W_{h_t}(t - t_j) = \frac{K_{h_t}(t - t_j)}{\sum_{j=1}^{J} K_{h_t}(t - t_j)}$ and $\sum_{j=1}^{J} W_{h_t}(t - t_j) = 1$. Estimator (3.4) is widely known as the Nadaraya-Watson (Nadaraya (1964), Watson (1964)) type kernel estimator. In smoothing estimation, Kernel works as a weighting function and bandwidth $h$ as a smoothing parameter. In equations (3.1) and (3.4), we used Epanechnikov Kernel and bandwidth is selected by the cross validation approach. Epanechnikov kernel is given by $K(u) = \frac{3}{4}(1 - u^2)1_{\{|u| \leq 1\}}$.

### 3.3. Bandwidth choices

The bandwidths of (3.2) and (3.4) may be selected either subjectively by examining the plots of the estimated parameter curves or using a data-driven bandwidth selection procedure. As demonstrated by the simulation studies in nonparametric estimation with two-step local polynomial estimators, such as
Fan and Zhang (2000), Wu et al. (2010) and Wu and Tian (2013a, b), subjective bandwidth choices obtained from examining the fitted curves of the estimators often produce appropriate bandwidths in real applications.

Two cross validation approaches, the “Leave-One-Subject-Out Cross Validation” (LSCV) and the “Leave-One-Time-Point-Out Cross Validation” (LTCV), have been proposed by Wu and Tian (2013a, b) for the selection of data-driven bandwidths under the unstructured nonparametric models. These cross validation approaches can be extended to the smoothing estimators (3.2) and (3.4) to provide a potential range of suitable bandwidths. Let
\[
\hat{\theta}^{(-i)}(t) \text{ with } 1 \leq i \leq n
\]
be the estimators (3.2) and (3.4) of \(\theta(t)\) computed using the sample with all the observations of the \(i^{th}\) subject deleted, and let \(w_i\) be a weight function which could be either \(1/(nm_i)\) or \(1/N\). The LSCV bandwidth \(h_{LSCV}\) is the minimizer of the LSCV score
\[
(3.5) \quad \text{LSCV}[y(\cdot)] = \sum_{j=1}^{J} \sum_{i \in S_j} w_i \{Y_i(t_j) - \hat{\theta}^{(-i)}(t_j)\}^2.
\]

For a heuristic justification of LSCV\([y(\cdot)]\), we can consider the expansion
\[
(3.6) \quad \text{LSCV}[y(\cdot)] = \sum_{j=1}^{J} \sum_{i \in S_j} w_i \{Y_i(t_j) - \theta(t_j)\}^2
\]
\[+ \sum_{j=1}^{J} \sum_{i \in S_j} w_i \{\theta(t_j) - \hat{\theta}^{(-i)}(t_j)\}^2
\]
\[+ 2 \sum_{j=1}^{J} \sum_{i \in S_j} w_i \{Y_i(t_j) - \theta(t_j)\} \times \{\theta(t_j) - \hat{\theta}^{(-i)}(t_j)\}.
\]

The first term at the right-hand side of (3.6) does not involve the smoothing estimator, hence, does not depend on the bandwidth. The expected value of the third term at the right-hand side of (3.6) is zero, since the observations of the \(i^{th}\) subject is not included in \(\hat{\theta}^{(-i)}(t_j)\). Thus, by minimizing LSCV\([y(\cdot)]\), the LSCV bandwidth \(h_{LSCV}\) approximately minimizes the second term at the right-hand side of (3.6), which is approximately the average squared error
\[
(3.7) \quad \text{ASE}[y(\cdot)] = \sum_{j=1}^{J} \sum_{i \in S_j} w_i \{\theta(t_j) - \hat{\theta}^{(-i)}(t_j)\}^2.
\]

A potential drawback for the LSCV approach is that the minimization of the LSCV score (3.5) is often computationally intensive, particularly when the number of subjects \(n\) is large, which hampers its application potential in real applications. Thus, it is usually more practical to consider the alternative of \(k\)-fold LSCV, which is computed by deleting the observations of \(k > 1\) subjects in the computation of \(\hat{\theta}^{(-i)}(t_j)\).
Instead of deleting the subjects one at a time, the LTCV procedure deletes the observations at the time design points \( t = \{t_1, \ldots, t_J\} \). When \( J \) is smaller than \( n \), the LTCV procedure may be computationally simpler than the LSCV procedure. Let \( \hat{\theta}^{(-j)}(t) \) with \( 1 \leq j \leq J \) be the estimators of (3.2) and (3.4) of \( \theta(t) \) computed using the sample with all the observations at the time point \( t_j \) deleted. Then the value of \( \hat{\theta}^{(-j)}(t) \) at time point \( t_j \) is \( \hat{\theta}^{(-j)}(t_j) \), and the LTCV score for \( \hat{\theta}(t) \) is

\[
\text{LTCV}[y(\cdot)] = \sum_{j=1}^{J} \sum_{i \in S_j} w_i \{Y_i(t_j) - \hat{\theta}^{(-j)}(t_j)\}^2.
\]

The LTCV bandwidth \( h_{\text{LTCV}} \) is the minimizer of \( \text{LTCV}[y(\cdot)] \). Similar to the \( k \)-fold alternative for the LSCV, the \( k \)-fold LTCV bandwidths, which are obtained by deleting \( k > 1 \) time points in \( t \) each time, may also be used in practical applications to reduce the computing complexity when \( J \) is large.

**Remark 3.1.** The simulation results of Wu and Tian (2013a, b) have shown that the LTCV approach may lead to appropriate bandwidths under the unstructured nonparametric models. Theoretical and practical properties of both the LSCV and LTCV bandwidths warrant substantial further investigation. In practice, the LSCV and LTCV bandwidths may only be used to provide a rough range of the appropriate bandwidths. The bandwidths for a actual dataset may be selected by evaluating the overall information from \( \text{LSCV}[y(\cdot)] \), \( \text{LTCV}[y(\cdot)] \), the scientific interpretations and the smoothness of the estimates.

### 3.4. Bootstrap pointwise confidence intervals

Since different asymptotic distributions of the smoothing estimators may be obtained depending on the cross sectional designs and whether and how fast \( m_j, j = 1, \ldots, J \), converge to infinity, statistical inferences based on the asymptotic approximations may not be an appropriate option in practice, and a widely used inference approach for nonparametric analysis is through the “resampling-subject” bootstrap suggested in Hoover et al. (1998). Under the current context, we can obtain a pointwise bootstrap confidence interval for \( \theta(t) \) 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 \( \{\hat{\theta}^b(t) : b = 1, \ldots, B\} \) using (3.2) and (3.4) for each of the bootstrap samples. The lower and upper boundaries of the \([100 \times (1-\alpha)]\)th empirical quantile bootstrap pointwise confidence interval of \( \hat{\theta}(t) \) are the empirical lower and upper \([100 \times (\alpha/2)]\)th percentiles based on the bootstrap estimators \( \{\hat{\theta}^b(t) : b = 1, \ldots, B\} \). Alternatively, if \( SD\{\hat{\theta}^b(t)\} \) is the empirical standard deviation of \( \{\hat{\theta}^b(t) : b = 1, \ldots, B\} \), the \([100 \times (1-\alpha)]\)th normally approximated bootstrap pointwise confidence interval of \( \hat{\theta}(t) \) is

\[
\hat{\theta}(t) \pm Z_{1-\alpha/2} \times SD\{\hat{\theta}^b(t)\},
\]
where $Z_{1-\alpha/2}$ is the $[100 \times (1-\alpha/2)]$th percentile of the standard normal distribution.

**3.5. Coverage probability of smoothing estimators**

In this section, we will derive the theoretical results on the coverage probability of the smoothing estimators of the bootstrap confidence interval. From (3.1) and (3.4), we see that local polynomial smoothing estimator and kernel smoothing estimator are expressed as a linear combinations of MLE of $\theta(t_j)$. Since $\hat{\theta}(t_j)$ is the MLE of $\theta(t_j)$, hence $\sqrt{n_j}(\hat{\theta}(t_j) - \theta(t_j)) \to N(0, \frac{1}{I(\theta_j)})$, where $I(\theta_j)$ is the Fisher information and for geometric distribution, it is $\frac{n_j}{\theta_j^2(1-\theta_j)}$. Let us denote the weight functions of (3.1) and (3.4) as $W_{1j} = W_{q,p+1}(t_j, t; h)$ and $W_{2j} = W_{hi}(t - t_j)$ respectively. Therefore, the variances of the estimators (3.1) and (3.4) are respectively given by $\sum_{j=1}^{J} \frac{W_{1j}^2}{I(\theta_j)}$ and $\sum_{j=1}^{J} \frac{W_{2j}^2}{I(\theta_j)}$. We know that MLE is asymptotically unbiased and hence the distribution of the local polynomial and kernel smoothing estimators are given by $(\hat{\theta}^{(q)}(t) - \theta^{(q)}(t)) \to N(0, A = \sum_{j=1}^{J} \frac{W_{1j}^2}{I(\theta_j)})$ and $(\hat{m}(t) - m(t)) \to N(0, B = \sum_{j=1}^{J} \frac{W_{2j}^2}{I(\theta_j)})$. By Hall (2013) (page-13), the theoretical bootstrap 95% confidence interval for local polynomial smoothing estimator and kernel smoothing estimator are respectively given by $(\hat{\theta}^{(q)}(t) - J^{-1/2}x_{.95}\hat{A}, \hat{\theta}^{(q)}(t) + J^{-1/2}x_{.95}\hat{A})$ and $(\hat{m}(t) - J^{-1/2}x_{.95}\hat{B}, \hat{m}(t) + J^{-1/2}x_{.95}\hat{B})$ where $P(|z| < x_{.95}) = \alpha$ with $z \sim N(0, 1)$. The corresponding coverage error is $P(\hat{\theta}^{(q)}(t) - J^{-1/2}x_{.95}\hat{A} \leq \theta^{(q)}(t) \leq \hat{\theta}^{(q)}(t) + J^{-1/2}x_{.95}\hat{A}) - .95 = P(|J^{1/2}((\hat{\theta}^{(q)}(t) - \theta^{(q)}(t))/A| \leq x_{.95}) - .95.

**4. Application to BDHS CW data**

We apply our methods to the BDHS (2011) data to estimate mean fecundabilities of the married women whose conception ages were between 14 and 49 years. Following the practical definition of age at conception (any age between onset of menstrual cycle and menopause), we round up the observed age of conception to the nearest integer with $J = 36$ distinct time design points $\{t_1, t_2, \ldots, t_{36}\} = \{14, 13, \ldots, 49\}$. Thus, for a given $1 \leq j \leq J = 36$ and $i \in S_j$, we denote $Y_i(t_j)$ as CW observation of the $i$th woman at conception age $t_j$. In our preliminary analysis, we have used the CW data at each age of conception to get the raw ML estimates of $\theta(t_j)$ from Geometric distribution. Applying the two-step local linear estimators of (3.2) and kernel estimators (3.4) to the observed data $\{Y_i(t_j), t_j; 1 \leq j \leq J, 1 \leq i \leq n_j\}$, we compute the smoothing estimators $\hat{\theta}(t)$ of $\theta(t)$ on the entire time design points $\{t_1, t_2, \ldots, t_{36}\} = \{14, 13, \ldots, 49\}$. In our analyses of the BDHS data, we found maximum age at first conception as 37th year.

Figure 1 shows the local linear smoothing estimates (lls) and kernel smoothing estimates (ks) of fecundability and their corresponding bootstrap pointwise 95% confidence interval based on $B = 100$ bootstrap replications for three different conception wait in months. A bandwidth of $h = 5$ and Epanechnikov kernel were used. This bandwidth was chosen by examining the LSCV and LTCV scores.
and the smoothness of the fitted plot. Figure 2 mimics the Fig. 1 except the conception wait (CW), which are considered as three different years. In both figures, overlapping CW are considered in part (b) and (c) to demonstrate that fecundability goes down when average CW goes up. In Fig. 1 of left and right panel, we see that raw estimates of fecundability lies between 0.50 and 0.90 whereas the smoothing fecundability lies between 0.60 and 0.70 and solid black line predicts fecundability at any time point. From the other panels of Figs. 1 and 2, we see
Figure 2. Raw estimators (scatter plots), smoothing estimators (solid curves), and bootstrap pointwise 95% confidence intervals (dashed curves, B = 100 bootstrap replications) of the age specific probabilities of conception by conception ages. (1a) to (1f): Estimators based on three different conception wait in years and two different smoothing methods with bandwidth $h = 5$.

similar type of results except decreasing fecundability. Therefore, fecundability should be reported by the highest CW of the data used to estimate fecundability.

Bootstrap confidence bands have been constructed to demonstrate that bandwidth choice is made correctly. A wrong choice of bandwidth produces smoothing estimator that would jump out the bootstrap confidence band. For example, a bandwidth of $h = 2.5$ would produce an smoothing estimator that intersect the bootstrap confidence band. In both figures, dots represent the raw estimates of
Table 2. ML estimate, local linear smoothing estimate, kernel smoothing estimate and the difference of smoothing estimate for the estimation of fecundability from BDHS 2011 samples of couples, who conceived within one year of marriage. The Epanechnikov kernel and the bandwidth $h = 5$ are used for the smoothing estimators.

| Age at First Conception | Maximum Likelihood Estimate | Kernel Smoothing Estimate | Local Linear Smoothing Estimate | Difference of Smoothing Estimate |
|-------------------------|-----------------------------|---------------------------|-------------------------------|---------------------------------|
| 14                      | 0.1752                      | 0.1646575                 | 0.1639815                     | 0.000676                        |
| 15                      | 0.1535                      | 0.163802                  | 0.1636487                     | 0.0001533                       |
| 16                      | 0.1614                      | 0.1634811                 | 0.1635313                     | −0.0000502                      |
| 17                      | 0.1691                      | 0.1634539                 | 0.1636271                     | −0.0001732                      |
| 18                      | 0.161                       | 0.1632983                 | 0.1639471                     | −0.0006488                      |
| 19                      | 0.1705                      | 0.1630045                 | 0.1645188                     | −0.0015143                      |
| 20                      | 0.1524                      | 0.1631657                 | 0.1653911                     | −0.0022254                      |
| 21                      | 0.1547                      | 0.164218                  | 0.1666393                     | −0.0024213                      |
| 22                      | 0.1847                      | 0.1657446                 | 0.1683719                     | −0.0026273                      |
| 23                      | 0.1585                      | 0.1670734                 | 0.1707383                     | −0.0036649                      |
| 24                      | 0.1746                      | 0.1682099                 | 0.1739395                     | −0.0057296                      |
| 25                      | 0.1522                      | 0.1694103                 | 0.1782405                     | −0.0088302                      |
| 26                      | 0.1827                      | 0.1699842                 | 0.1839835                     | −0.0139993                      |
| 27                      | 0.1829                      | 0.1692527                 | 0.1916006                     | −0.0223479                      |
| 28                      | 0.1803                      | 0.1699002                 | 0.2016228                     | −0.0317226                      |
| 29                      | 0.119                       | 0.179044                  | 0.2146834                     | −0.0356394                      |
| 30                      | 0.1379                      | 0.2021976                 | 0.2315132                     | −0.0293156                      |
| 31                      | 0.1852                      | 0.2354042                 | 0.2529279                     | −0.0175237                      |
| 32                      | 0.5                         | 0.2670875                 | 0.2798081                     | −0.0127206                      |
| 33                      | 0.2222                      | 0.287872                  | 0.3130741                     | −0.0252021                      |

Fecundability (ML estimates), solid black line represents smoothing estimates of fecundability and dotted line represents the 95% pointwise bootstrap confidence band. By looking at the figures, we can say that kernel smoothing estimator is little rough compare to local linear smoothing estimator. Table 2 show the raw and smoothing estimates of fecundability and the difference of smoothing estimates of fecundability within one year of marriage. In Table 2, we see that kernel smoothing estimators have consistently lower smoothing estimates compared to the local linear smoothing estimates except at lower boundary ages 14 and 15. As mean natural fecundability is estimated from couples who do not use contraception before giving first birth, hence most of the first conception occur before woman reaches mid to late thirties. Fecundability is found to be little high in some cases, which could be explained by three reasons: (1) in some conception ages (age bin), number of subject could be small and these subjects also have smaller CW, which leads to higher estimate of fecundability (2) Raw fecundability is estimated from geometric model, which goes down in harmonic nature as the CW goes up. (3) Rounding approach might reduce the CW. These estimates of fecundability is consistent with the mean fecundability of table 8 of
Potter and Parker (1964) corresponding to conception delay of six months (Fig. 1) and twelve months (Fig. 2 and Table 2). However, Potter and Parker (1964) estimated fecundability by considering all women as an homogeneous cohort whereas we estimated fecundability by considering women on different age-specific cohort (ages as first conception) and our methods can predict fecundability at any point on entire time design points.

5. Simulation study

Following the data structure of Subsection 2.1, we generate in each sample \( n = 2500 \) subjects with 100 subjects randomly allotted to each of the 25 ages in \( \{t_1, t_2, \ldots, t_{25}\} = \{1, 2, \ldots, 25\} \). Each of these subjects has conception wait randomly generated from the geometric model with probability of conception as \( p = .10, p = .20, p = .30, p = .40, p = .50 \) and \( p = .60 \). We repeated the above sampling procedure for \( S = 500 \) times and computed monthly probability of conception for each of the 25 ages for each of the simulated samples. We then

| Age | \( P = .10 \) | \( P = .20 \) | \( P = .30 \) | \( P = .40 \) | \( P = .50 \) | \( P = .60 \) |
|-----|---------------|---------------|---------------|---------------|---------------|---------------|
| 1   | 1.1054        | 1.1320        | 1.1476        | 1.1223        | 1.1645        | 1.1530        |
| 2   | 1.0237        | 1.0125        | 1.0575        | 1.0507        | 1.0685        | 1.0147        |
| 3   | 0.9086        | 0.8930        | 0.9241        | 0.9127        | 0.9165        | 0.8803        |
| 4   | 0.7661        | 0.7760        | 0.7776        | 0.7365        | 0.7500        | 0.7541        |
| 5   | 0.6489        | 0.6709        | 0.6509        | 0.5984        | 0.6236        | 0.6398        |
| 6   | 0.5771        | 0.5790        | 0.5555        | 0.5259        | 0.5382        | 0.5487        |
| 7   | 0.5371        | 0.4991        | 0.4983        | 0.5025        | 0.4784        | 0.4873        |
| 8   | 0.5071        | 0.4376        | 0.4738        | 0.4911        | 0.4457        | 0.4486        |
| 9   | 0.4768        | 0.4058        | 0.4615        | 0.4682        | 0.4354        | 0.4221        |
| 10  | 0.4550        | 0.4051        | 0.4429        | 0.4423        | 0.4177        | 0.3982        |
| 11  | 0.4460        | 0.4261        | 0.4241        | 0.4316        | 0.3847        | 0.3773        |
| 12  | 0.4410        | 0.4516        | 0.4180        | 0.4432        | 0.3626        | 0.3741        |
| 13  | 0.4331        | 0.4672        | 0.4159        | 0.4577        | 0.3644        | 0.3934        |
| 14  | 0.4267        | 0.4643        | 0.4040        | 0.4497        | 0.3878        | 0.4182        |
| 15  | 0.4274        | 0.4441        | 0.3953        | 0.4303        | 0.4168        | 0.4346        |
| 16  | 0.4405        | 0.4329        | 0.4126        | 0.4214        | 0.4290        | 0.4491        |
| 17  | 0.4634        | 0.4489        | 0.4657        | 0.4366        | 0.4345        | 0.4676        |
| 18  | 0.4887        | 0.4873        | 0.5476        | 0.4924        | 0.4596        | 0.5065        |
| 19  | 0.5357        | 0.5495        | 0.6357        | 0.5964        | 0.5097        | 0.5885        |
| 20  | 0.6408        | 0.6512        | 0.7235        | 0.7236        | 0.5968        | 0.7100        |
| 21  | 0.8235        | 0.8098        | 0.8464        | 0.8474        | 0.7544        | 0.8640        |
| 22  | 1.0662        | 1.0244        | 1.0113        | 0.9913        | 0.9870        | 1.0477        |
| 23  | 1.3133        | 1.2210        | 1.1758        | 1.1788        | 1.2315        | 1.2377        |
| 24  | 1.5110        | 1.3427        | 1.3229        | 1.3847        | 1.4279        | 1.4184        |
| 25  | 1.6663        | 1.4477        | 1.4683        | 1.5744        | 1.5847        | 1.6039        |
apply the local linear smoothing estimators and kernel smoothing estimators on these 25 raw estimates of fecundability for each of the 500 simulated samples. We then computed average Bias, MSE and Coverage probability for each of the smoothing estimators. We have also computed the relative MSE, which is defined as the ratio of MSEs computed by local linear smoothing estimator to kernel smoothing estimators. In Table 3, we have presented only the ratio of the MSEs for various conception probabilities. In the tabular presentation, we ignored the bias differences between two smoothing estimators and coverage probability, because average biases produced by two smoothing estimators from 500 simulated samples are very small and the coverage probabilities are around 95% at all 25 ages for two smoothing estimators. From Table 3, we see that MSEs obtained from kernel smoothing estimators are smaller at few points of the bounder of the support and in all other points, the MSEs from kernel smoothing estimates are larger in all six conception probabilities, which indicates the superiority of local linear smoothing estimator over kernel smoothing estimator.

6. Asymptotic results

We establish in this section the asymptotic bias, variance and mean squared errors (MSE) of the local polynomial smoothing estimator (3.1) and kernel smoothing estimator (3.4) and their relationship. The asymptotic properties of raw estimator $\hat{\theta}(t)$ can be derived by applying the properties of MLE.

6.1. Asymptotic properties of the raw estimators

Following Subsection 3.1, $\theta(t_j)$ at each of the time design point $t_j \in t$ is estimated by the MLE $\tilde{\theta}(t_j)$. Suppose that the classical regularity conditions of the MLEs, i.e., the conditions of Theorem 5.41 of van der Vaart (1998), are satisfied. Then, for all $t_j \in t$, $n_j^{1/2} \left[ \theta(t_j) - \hat{\theta}(t_j) \right]$ has asymptotically the $\mathcal{N}(0, I^{-1}(\theta(t_j)))$ distribution, where $I(\theta(t_j))$ is the Fisher information matrix at $\theta(t_j)$. It follows that $\tilde{\theta}(t_j)$ is asymptotically unbiased for $\theta(t_j)$, i.e., $\mathbb{E} \left[ \tilde{\theta}(t_j) \right] \equiv \theta(t_j)$ and the asymptotic variance of $\tilde{\theta}(t_j)$ is $n_j^{-1} I^{-1}(\theta(t_j))$. At different time points $t_j \neq t_k$, $\tilde{\theta}(t_j)$ and $\tilde{\theta}(t_k)$ are uncorrelated, and the covariance $\text{Cov}[\tilde{\theta}(t_j), \tilde{\theta}(t_k)] = 0$.

6.2. Asymptotic properties for local polynomial smoothing estimators

We assume the following asymptotic assumptions for the two-step local polynomial estimators $\hat{\theta}^{(p)}(t)$ given in (3.1):

A1. If $n \to \infty$, then $h \to 0$, $n^{1/2} h^{p-q+1} \to \infty$, $Jh \to \infty$, and $n J h^{2q+1} \to \infty$.

A2. The design time points $\{t_1, t_2, \ldots, t_J\}$ are independent and identically distributed with density function $g(t)$. For all $1 \leq j \leq J$.

A3. $\theta(t)$ are $p+1$ times continuously differentiable with respect to $t$.

A4. The kernel function $K(\cdot)$ is a bounded symmetric probability density function with support within a bounded set $[-a, a]$ for some $a > 0$.

Let $K_{d,p+1}(t) = c_{d,p+1} S^{-1} (1, t, \ldots, t^p)^T K(t)$ be the equivalent kernel of local polynomial fit with $S = (s_{kl})_{k,l=0,1,\ldots,p}$ and $s_{kl} = \int K(u) u^{k+l} du$, $B_{p+1}(K) = \int K(u) u^{p+1} du$ and $V(K) = \int K^2(u) du$. The following theorem summarizes the
asymptotic expressions of the bias, variance and mean squared errors of the local polynomial estimator of the \(q\)th derivative of \(\theta(t)\) based on the raw estimates of \(\hat{\theta}(t_j), j = 1, \ldots, J\).

**Theorem 1.** Suppose that the Assumptions A1–A4 are satisfied. When \(n\) is sufficiently large,

\[
\text{Bias}\{\hat{\theta}^{(q)}(t)\} = E\{\hat{\theta}^{(q)}(t)\} - \theta^{(q)}(t) = \frac{q!h^{p-q+1}}{(p+1)!}\theta^{(p+1)}(t)B_{p+1}(K_{q,p+1})[1 + o_p(1)]
\]

\[
\text{Var}\{\hat{\theta}^{(q)}(t)\} = \frac{(q!)^2}{Jh^{2q+1}g(t)}V(K_{q,p+1}) \times n_j^{-1}I^{-1}[\theta(t_j)][1 + o_p(1)]
\]

and the asymptotic expression of the mean squared error (MSE)

\[
\text{MSE}\{\hat{\theta}^{(q)}(t)\} = \text{Bias}\{\hat{\theta}^{(q)}(t)\}^2 + \text{Var}\{\hat{\theta}^{(q)}(t)\},
\]

**Proof.** See Appendix A2.

**Remark 6.1.** Special cases of Theorem 1 can be easily derived from (6.1), (6.2) and (6.3). For the local linear estimator of \(\theta(t)\), we have \(q = 0\) and \(p = 1\), so that the asymptotic MSE of \(\hat{\theta}(t)\) is

\[
\text{MSE}\{\hat{\theta}(t)\} = \{h^4B^2(t) + h^{-1}(nJ)^{-1}V(t)\}[1 + o_p(1)],
\]

where \(B(t) = \theta''[t]B_2(K_{0,2})/2\) and

\[
V(t) = [ng(t)]^{-1}V(K_{0,2})I^{-1}[\theta(t)].
\]

Setting \(\partial\text{MSE}\{\hat{\theta}(t)\}/\partial h\) to zero, the theoretically optimal bandwidth \(h_{opt}\) which minimizes the dominating term at the right side of (6.4) is

\[
h_{opt} = (nJ)^{-1/5}[V(t)]^{1/5}[4B^2(t)]^{-1/5}.
\]

Substituting \(h_{opt}\) into (6.4), the MSE of the local linear estimator \(\hat{\theta}(t)\) is

\[
\text{MSE}\{\hat{\theta}(t)\} = (nJ)^{-4/5}[V(t)]^{4/5}[B(t)]^{2/5}(2^{-8/5} + 2^{2/5}),
\]

which suggests that the optimal rate for the MSE of \(\hat{\theta}(t)\) to converge to zero is \((nJ)^{4/5}\).

### 6.3. Asymptotic properties for kernel smoothing estimators

We assume the following asymptotic assumptions for the two-step kernel smoothing estimators given in (3.4):
A1. If \( h \to \infty \), then \( \frac{h^v}{m} \to 0 \) when \( v \to \infty \). This assumption states that \( v \) goes to infinity in a faster rate than \( h \).

A2. The design time points \( \{t_1, t_2, \ldots, t_J\} \) are independent and identically distributed with density function \( f(t) \). For all \( 1 \leq j \leq J \).

A3. \( f(t) \) are \( v + 1 \) times continuously differentiable with respect to \( t \).

A4. The kernel function \( K(\cdot) \) is a bounded symmetric probability density function with support within a bounded set \([-a, a]\) for some \( a > 0 \).

A5. The kernel function \( K(\cdot) \) is second order kernel i.e., \( k_j(k) = 0 \); for \( j < 2 \) where \( k_j(k) = \int_{-\infty}^{\infty} u^j k(u) du \).

Suppose \( R(k) = \int_{-\infty}^{\infty} k^2(u) du \). Bias, Variance and MSE (Mean Squared Error) of the kernel smoothing estimator \( \hat{m}(t) \) of (3.4) can be obtained by separately finding the mean and variance of the numerator and denominator of (3.4). From Kernel density estimator, we know that \( \hat{f}(t) = \frac{1}{Jh} \sum_{j=1}^{J} k\left(\frac{t_j - t}{h}\right) \).

\[
E(\hat{f}(t)) = \frac{1}{h} k\left(\frac{t_j - t}{h}\right)
= \int_{-\infty}^{\infty} \frac{1}{h} k\left(\frac{z - t}{h}\right) f(z)
= \int_{-\infty}^{\infty} k(u) f(t + hu) du
= f(t) + \frac{1}{2!} f^{(2)}(t) h^2 k_2(k) + O(h^4)
\]

If kernel is of \( v \)th order then \( E(\hat{f}(t)) = f(t) + \frac{1}{v!} f^{(v)}(t) h^v k_v(k) + O(h^v) \). Since \( k(\cdot) \) is an iid and also a linear estimator, so,

\[
\text{Var}(\hat{f}(t)) = \frac{1}{Jh^2} \text{Var}\left(K \left(\frac{t_j - t}{h}\right)\right)
= \frac{1}{Jh^2} \left( \frac{1}{h} E K \left(\frac{t_j - t}{h}\right) \right)^2
= \frac{1}{Jh^2} \left( \frac{1}{h} E K \left(\frac{t_j - t}{h}\right) \right)^2
= \frac{1}{Jh} \int_{-\infty}^{\infty} K(u) f(t + hu) du
= \frac{f(t) R_{k}(k)}{Jh} + O(h).
\]

The second term of second equality is equivalent to \( f(t) \) plus a tiny bias term, which vanishes as \( T \to \infty \). Now we have to find the mean and variance of the numerator.

\[
E \left\{ \int \theta \hat{f}(t, \theta) d\theta \right\} = E \left\{ \frac{1}{h} k \left(\frac{t_j - t}{h}\right) \theta_j \right\}
= \frac{1}{h} \int \int k \left(\frac{z - t}{h}\right) \theta f(z, \theta) dz d\theta
\]
\[
\begin{align*}
&= \int \int k(u) \theta f(uh + t, \theta) dud\theta \\
&= \int k(u)f(uh + t)du \int \theta f(\theta | uh + t) d\theta \\
&= \int k(u)f(uh + t)m(uh + t)du \\
&= f(t)m(t) + h^2 k_2(k) \left[ f^{(1)}(t)m^{(1)}(t) + \frac{f^{(2)}(t)m(t)}{2!} \right. \\
&\quad \left. + \frac{f(t)m^{(2)}(t)}{2!} + 0(h^2) \right].
\end{align*}
\]

\[
(6.9) \quad \text{Var} \left\{ \int \theta \hat{f}(t, \theta) d\theta \right\} \\
= \frac{1}{Jh^2} E \left\{ K \left( \frac{t_j - t}{h} \right) \theta_j \right\}^2 - 0 \left( \frac{1}{J} \right) \\
= \frac{1}{Jh} \int \int \theta^2 K^2(u)f(uh + t, \theta) dud\theta \\
= \frac{1}{Jh} \int k^2(u)f(uh + t)du \int \theta^2 f(\theta | uh + t) d\theta \\
= \frac{1}{Jh} \int k^2(u)f(uh + t)du \left[ \text{Var}(\theta | uh + t) + \{E(\theta | uh + t)\}^2 \right] \\
= \frac{1}{Jh} \int K^2(u)f(uh + t)\{\sigma^2(uh + t) + m^2(uh + t)\}du \\
= \frac{1}{Jh} R(k)f(t)[\sigma^2(t) + m^2(t)].
\]

\[
(6.10) \quad \text{Cov} \left\{ \hat{f}(t), \int \theta \hat{f}(t, \theta) d\theta \right\} = \frac{1}{J^2h^2} E \left\{ \sum_{j=1}^J k^2 \left( \frac{t_j - t}{h} \right) \theta_j \right\} - 0 \left( \frac{1}{J} \right) \\
= \frac{1}{Jh} \int \int \theta K^2(u)f(uh + t, \theta) dud\theta \\
= \frac{1}{Jh} \int k^2(u)f(uh + t)du \int \theta f(\theta | uh + t) d\theta \\
= \frac{1}{Jh} \int k^2(u)f(uh + t)m(uh + t)du \\
= \frac{1}{Jh} R(k)f(t)m(t).
\]

Last terms of equations (3.8) through (6.4) are obtained by transformation of variable \( u = \frac{z - t}{h} \), using above assumptions and applying Taylor series on \( f(uh + t), m(uh + t), m^2(uh + t) \) and \( \sigma^2(uh + t) \) in the neighbourhood of \( uh \). The following theorem summarizes the asymptotic expressions of the bias, variance and mean squared errors of the kernel smoothing estimator.
THEOREM 2. Suppose that the assumptions A1–A5 are satisfied. When $n$ is sufficiently large,

\begin{equation}
(6.11) \quad \text{Bias}\{\hat{m}(t)\} = E\{\hat{m}(t)\} - m(t) \approx \frac{h^2k_2(k)}{2} \left[ m^{(2)}(t) + \frac{2f^{(1)}(t)m^{(1)}(t)}{f(t)} \right]
\end{equation}

\begin{equation}
(6.12) \quad \text{Var}\{\hat{m}(t)\} \approx \frac{R(k)K_2(k)}{Jh_f(t)}
\end{equation}

and the asymptotic expression of the mean squared error (MSE)

\begin{equation}
(6.13) \quad \text{MSE}\{\hat{m}(t)\} \approx \text{Bias}\{\hat{m}(t)\}^2 + \text{Var}\{\hat{m}(t)\}.
\end{equation}

Setting $\frac{\partial}{\partial h}\text{MSE}\{\hat{m}(t)\}$ to zero, the theoretical optimal bandwidth $h_{opt}$, which minimizes right side of equation (6.7) is $h_{opt} = \{\frac{R(k)}{Jf(t)k_2(t)}\}^{1/5}A^{-2/5}$ where $A = m^{(2)}(t) + \frac{2f^{(1)}(t)m^{(1)}(t)}{f(t)}$.

**Proof.** See Appendix A3.

THEOREM 3. The local polynomial smoothing estimator $\hat{\theta}^{(q)}(t)$ and kernel smoothing estimator $\hat{m}(t)$ have the following relationship.

\begin{equation}
(6.14) \quad \hat{\theta}^{(q)}(t) = \frac{D_1D_2D_4}{D_3} \hat{m}(t).
\end{equation}

**Proof.** See Appendix A4.

Let $\text{MSE}_l$ and $\text{MSE}_k$ are the MSEs of local polynomial smoothing estimators and kernel smoothing estimators respectively. Now, $\text{MSE}_l = h^{2p-2q+2}B_1 + \frac{1}{k^2q+1}B_2$ where $B_1 = \{\frac{q!}{(p+1)!}\}^2 B_{p+1}(K_{q,p+1})^2$ and $B_2 = \{\frac{q!}{(p+1)!}\}^2 V(K_{q,p+1}) n_j^{-1} \times I^{-1}[\theta(t_j)]$. So, we have $h = \frac{h^{2p+3}B_1 + B_2}{k^2q\text{MSE}_l}$. Similarly, we have $h = \frac{h^3A_1 + A_2}{\text{MSE}_k}$ for kernel smoothing estimator where $A_1 = \{\frac{k_2(k)}{2}[m^{(2)}(t) + 2\frac{f^{(1)}(t)m^{(1)}(t)}{f(t)}]\}^2$ and $A_2 = \frac{R(k)K_2(k)}{Jf(t)}$. Equating for $h$, obtained from above two MSEs, we have the following asymptotic result.

THEOREM 4. When $n_j \to \infty$, $q = 0$ and $p = 1$ the ratio of the MSEs are given as

\begin{equation}
(6.15) \quad \frac{\text{MSE}_l}{\text{MSE}_k} = \frac{\theta^4(t_j)}{M + Qk_2(k)}.
\end{equation}

Where $Q = \frac{4R(k)}{Jf(t)h^3k_2(k)}$ and $M = \{[m^{(2)}(t) + 2\frac{f^{(1)}(t)m^{(1)}(t)}{f(t)}]\}^2$.

**Proof.** See Appendix A5.

Remark 6.2. In the simulation results, we discussed the conditions under which the above ratio is less than 1 and vice-versa.
7. Discussion

We proposed a class of time-varying parametric models for smoothing estimation of the parameter by a two-step local polynomial smoothing estimator and by a kernel smoothing estimator with cross-sectional data. These models, which belong to a class of structural nonparametric models, are useful for studies with large sample sizes. Asymptotic properties of the raw estimators and smoothing estimators are also derived. MSE from kernel smoothing estimator is expressed in terms of MSE from local polynomial smoothing estimator. Mathematical relationship between two smoothing estimator has been established by equation (6.14). We show the application of our procedure by a real survey data. Simulation study is done to check finite sample properties as well as to show the superiority of the local polynomial smoothing estimator over kernel smoothing estimator. Parameters from time-variant parametric regression models and semi-parametric regression models can also be smoothed by these smoothing estimators. Odds Ratio from logistic model, hazard ratio from Cox model and competing risk model and C-statistics from any regression model can also be smoothed by these smoothing techniques when time-variant data are available. The theoretical results presented here are applicable to cross-sectional data, which could be extended to longitudinal data by adding covariance structure to these theoretical results. Application and simulation under longitudinal framework has not been considered because of redundancy and also to keep the size of paper concise.

There are a number of theoretical and methodological aspects that warrant further investigation. Theoretical and simulation 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.

If the size of the sample is not large enough at each time point, we can apply the one-step kernel log-likelihood to fit the model (2.2). This would entail choosing the parameters of a local polynomial to minimize

\[
\sum_{j=1}^{n} \left\{ \log \left( \sum_{q=0}^{p} \beta_q(t)(t_j - t)^q \right) \right. \\
+ \left. (y_j - 1) \log \left(1 - \sum_{q=0}^{p} \beta_q(t)(t_j - t)^q\right) \right\} K_h(t_j - t)
\]

where summation is over all the data points. Under current framework, considering one-step kernel log likelihood estimation will tremendously increase the size of the paper and hence out of scope of the current paper. But in our future research, we will consider a detail study between one-step smoothing with two-step smoothing under cross-sectional and longitudinal framework.
Appendix A: Proof of theoretical results

A.1 Useful approximation for the equivalent kernels

The following approximations for the equivalent kernel function $W_{q,p+1}(t_j, t; h)$ are used in computing the asymptotic bias and variance of $\hat{\theta}^{(q)}(t)$:

(A.1) \[ W_{q,p+1}(t_j, t; h) = \frac{q!}{J h^{q+1} g(t)} K_{q,p+1} \left( \frac{t_j - t}{h} \right) [1 + o_p(1)], \quad j = 1, \ldots, J; \]

(A.2) \[ \sum_{j=1}^{J} W_{q,p+1}(t_j, t; h)(t_j - t)^k = q! l_{[k=q]}, \quad k = 0, 1, \ldots, p; \]

(A.3) \[ \sum_{j=1}^{J} W_{q,p+1}(t_j, t; h)(t_j - t)^{p+1} = q! h^{p-q+1} B_{p+1}(K_{q,p+1})[1 + o_p(1)], \quad k = 1, \ldots, p; \]

(A.4) \[ \sum_{j=1}^{J} W_{q,p+1}^2(t_j, t; h) = \frac{(q!)^2}{J h^{2q+1} g(t)} V(K_{q,p+1})[1 + o_p(1)], \]

where $K_{q,p+1}(t)$, $B_{p+1}(K)$ and $V(K)$ are defined in Theorem 1. Proofs of equations (A.1)–(A.4) are given Fan and Zhang (2000, Appendix A, Lemma 1 and Lamma 2).

A.2 Proof of Theorem 1

By equation (3.1), asymptotic unbiased property of MLE, equations (A.1) to (A.4) and Taylor expansion on $\theta(t_j)$, we have

\[
E\{\hat{\theta}^{(q)}(t)\} \approx \sum_{j=1}^{J} W_{q,p+1}(t_j, t; h)\theta(t_j)
\]

\[
= \sum_{j=1}^{J} W_{q,p+1}(t_j, t; h)
\]

\[
\times \left[ \sum_{k=0}^{p} \theta^{(k)}(t_j) \frac{(t_j - t)^k}{k!} + \frac{\theta^{(p+1)}(t_j)(t_j - t)^{p+1}}{(p+1)!} + O_p(1) \right]
\]

\[
= \theta^{(q)}(t) + \frac{q! \theta^{p+1}(t_j) h^{p-q+1}}{(p+1)!} \times B_{p+1}(K_{q,p+1})[1 + O_p(1)]
\]

\[
\text{Var}\{\hat{\theta}^{(q)}(t)\} = \text{Var}\left\{ \sum_{j=1}^{J} W_{q,p+1}(t_j, t; h)\tilde{\theta}(t_j) \right\}
\]

\[
= E\left\{ \sum_{j=1}^{J} W_{q,p+1}(t_j, t; h)\tilde{\theta}(t_j) - \sum_{j=1}^{J} W_{q,p+1}(t_j, t; h)\theta(t_j) \right\}^2
\]

\[
= \sum_{j=1}^{J} W_{q,p+1}^2(t_j, t; h) E[\tilde{\theta}(t_j) - \theta(t_j)]^2
\]
A.3 Proof of Theorem 2

Let $A = \int \theta f[t, \theta(t)]d\theta$ and $B = \hat{f}(t)$. Bias, Variance and MSE of the kernel smoothing estimator $\hat{m}(t) = \frac{A}{B}$ can be obtained by using the approximation formula for the ratio of two random variables as follows:

$$E(\hat{m}(t)) \approx \frac{EA}{EB}$$

$$f(t)m(t) + h^2 k_2(k) \left[ \frac{f^{(1)}(t)m^{(1)}(t)}{f(t)} + \frac{f^{(2)}(t)m(t)}{2!f(t)} + \frac{f(t)m^{(2)}(t)}{2!f(t)} \right]$$

$$m(t) + h^2 k_2(k) \left[ \frac{f^{(1)}(t)m^{(1)}(t)}{f(t)} + \frac{f^{(2)}(t)m(t)}{2!f(t)} + \frac{f(t)m^{(2)}(t)}{2!f(t)} \right]$$

$$= m(t) + h^2 C_1$$

$$= [m(t) + h^2 C_1][1 + h^2 C_2]^{-1}$$

$$= m(t) - m(t)h^2 C_2 + h^2 C_1 - h^4 C_1 C_2$$

$$= m(t) - \frac{m(t)h^2 f^{(2)}(t)k_2(k)}{2f(t)}$$

$$+ h^2 k_2(k) \left[ \frac{f^{(1)}(t)m^{(1)}(t)}{f(t)} + \frac{f^{(2)}(t)m(t)}{2!f(t)} + \frac{m^{(2)}(t)}{2} \right]$$

$$- h^4 k_2(k) \left[ \frac{f^{(1)}(t)m^{(1)}(t)}{f(t)} + \frac{f^{(2)}(t)m(t)}{2!f(t)} + \frac{m^{(2)}(t)}{2} \right] \left[ \frac{f^{(2)}(t)k_2(k)}{2f(t)} \right]$$

$$= m(t) + \frac{h^2 k_2(k)}{2} \left[ m^{(2)}(t) + 2 \frac{f^{(1)}(t)m^{(1)}(t)}{f(t)} \right] - 0(h^4)$$

$$\text{Var} \left( \frac{A}{B} \right) \approx \left( \frac{EA}{EB} \right)^2 \left[ \text{Var}(A) + \frac{\text{Var}(B)}{(EB)^2} - \frac{2\text{Cov}(A, B)}{(EA)(EB)} \right]$$

$$= \frac{1}{(EB)^2} \left[ \text{Var}(A) + \text{Var}(B)\hat{m}^2(t) - 2\text{Cov}(A, B)\hat{m}(t) \right]$$

$$= \frac{1}{\{f(t)\}^2} \left[ R(k)f(t)Th(k_2(k) + \hat{m}^2(t)) + \frac{R(k)f(t)Th}{k_2(k) + \hat{m}^2(t)} \right]$$
\[
\frac{R(k)K_2(k)}{Thf(t)} - 2\frac{R(t)f(t)\hat{m}(t)}{Th}\]

A.4 Proof of Theorem 3

From equation (3.1), we have

\[h = D_1D_2\hat{\theta}q(t),\]

where

\[D_1 = q!e_q+1,p+1[T_p^T(t)G(t;h)T_p(t)]^{-1}\]

and

\[D_2 = \sum_{j=1}^J T_j^T(t_0,\ldots,K(t_j-h),\ldots,0)\hat{\theta}(t_j).\]

From equation (3.4), we have

\[h = D_3D_4\hat{m}(t),\]

where

\[D_3 = \sum_{j=1}^J K(t_j-h)\hat{\theta}(t_j),\]

\[D_4 = \sum_{j=1}^J K_h(t-t_j).\]

Equating for \(h\), we get the relationship between kernel smoothing estimator and local polynomial smoothing estimator as a finite sample property.

A.5 Proof of Theorem 3

\[
\text{MSE}_l = \frac{h^{2p+3}B_1 + B_2}{h^{2q}(h^5A_1 + A_2)}
\]

\[
= \frac{h^{2p+3}\left\{\frac{(q!\theta^{q+1}(t_j)}{(p+1)!} \times B_{p+1}(K_q,p+1)\right\}^2 + \frac{(q!)^2}{Jg(t)} V(K_q,p+1)n_j^{-1}I^{-1}[\theta(t_j)]}{h^{2q} \left\{\frac{k_2(k)}{2} \left[m^{(2)}(t) + 2\frac{f^{(1)}(t)m^{(1)}(t)}{f(t)}\right]\right\}^2 + \frac{R(k)K_2(k)}{Jf(t)}}
\]

For local linear estimator, \(q = 0\), \(p = 1\), \(k_2(k) = \int u^2k(u)du = B_2(k_0,2)\) and \(V(k_0,2) = R(k)\), we have

\[
= \frac{h^5\theta^{4}(t_j)k_2^2(k)}{4} + \frac{1}{Jf(t)} R(k)n_j^{-1}I^{-1}[\theta(t_j)]
\]

\[
= \frac{h^5k_2^2(k)}{4} M + \frac{R(k)k_2(k)}{Jf(t)}
\]

\[
\theta^{4}(t_j) + \frac{4R(k)n_j^{-1}I^{-1}[\theta(t_j)]}{Jf(t)h^5k_2^2(k)}
\]

\[
= \frac{M + 4R(k)}{Jf(t)h^5k_2^2(k)}
\]

\[
\theta^{4}(t_j) + \frac{Q}{n_jI[\theta(t_j)]}
\]

\[
= \frac{\theta^{4}(t_j)}{M + Qk_2(k)}
\]

when \(n_j \to \infty\)
References

BDHS (2011). Bangladesh Demographic and Health Survey.

Bongaarts, J. (1975). A method for the estimation of fecundability, *Demography*, 12(4), 645–660.

Chowdhury, M. and Umbach, D. (2012). Some Bayesian analyses of fecundability, *Pak. J. Statist.*, 28(3), 293–302.

Chowdhury, M., Wu, C. and Modarres, R. (2017). Local Box-Cox transformation on time varying parametric models for smoothing estimation of conditional CDF with longitudinal data, *Journal of Statistical Computation and Simulation*, 87(15), 2900–2914.

Chowdhury, M., Wu, C. and Modarres, R. (2017). Nonparametric estimation of conditional distribution function with longitudinal models and time-varying parametric models, https://doi.org/10.1007/s00184-017-0634-z.

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, *J. R. Statist. Soc. Ser. B*, 62, 303–322.

Hall, P. (2013). *The Bootstrap and Edgeworth Expansion*, Springer.

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.

Jain, A. K. (1969). Fecundability and its relation to age in a sample of Taiwanese women, *Population Studies*, 23, 69–85.

Majumdar, H. and Sheps, M. C. (1970). Estimators of a type 1 geometric distribution from observation on conception times, *Demography*, 7, 349–360.

Nadaraya, E. A. (1964). On estimating regression, *Theory Probab. Appl.*, 9(1), 141–142.

Pandey, H. and Yadav, A. K. (2013). A probability model for first conception and its Bayesian analysis, *J. Comp. & Math. Sci.*, 4(3), 179–185.

Perrin, E. B. and Sheps, M. C. (1964). Human reproduction: A stochastic process, *Biometrics*, 20(1), 28–45.

Potter, R. G. and Parker, M. P. (1964). Predicting time required to conceive, *Population Studies*, 18(1), 99–116.

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

Sheps, M. C. and Menken, J. A. (1973). *Mathematical Models of Conception and Birth*, University of Chicago Press.

van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge, UK.

Watson, G. S. (1964). Smooth regression analysis, *The Indian Journal of Statistics, Series A*, 26(4), 359–372.

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, *J. Am. Stat. Assoc.*, 108(503), 971–982.

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