Estimation in Covariate Adjusted Regression

Damla Şentürk[1] and Danh V. Nguyen[2]
[1]Department of Statistics, Pennsylvania State University
University Park, PA 16802, U.S.A.
email: dsenturk@stat.psu.edu
and
[2]Department of Public Health Sciences,
University of California School of Medicine,
Davis, CA 95616, U.S.A.
email: ucdnguyen@ucdavis.edu

Abstract: The method of covariate adjusted regression was recently proposed for situations where both predictors and response in a regression model are not directly observed, but are observed after being contaminated by unknown functions of a common observable confounder in a multiplicative fashion. One example is data collected for a study on diabetes, where the variables of interest, systolic and diastolic blood pressures and glycosolated hemoglobin levels are known to be influenced by an observable confounder, body mass index. An estimation procedure based on equidistant binning (EB), currently available, gives consistent estimators for the regression coefficients adjusted for the confounder. In this paper, we propose two new estimation procedures based on nearest neighbor binning (NB) and local polynomial modeling (LP). Even though, the three methods perform similarly in terms of their bias, it is shown through simulation studies that NB has smaller variance compared to EB, and LP yields substantially lower variance relative to the two binning methods for small to moderate sample sizes. The consistency and convergence rates of the proposed estimators of LP, with the smallest MSE, are also established. We illustrate the proposed method of LP with the above mentioned diabetes data, where the goal is to uncover the regression relation between the response, glycosolated hemoglobin levels, and the predictors, systolic and diastolic blood pressures, adjusted for body mass index.

Keywords: Equidistant binning; Local polynomial regression; Multiplicative effects; Nearest neighbor binning; Smoothing; Varying-coefficient models.

Running Title: Estimation in CAR

1. Introduction

In many medical studies measurements are taken on potential confounding variables that are believed to affect the primary variables of interest. In such studies, a common interest is to uncover the relationships between the primary variables of interest adjusted for the effects of the confounders. One example is data collected for a study on diabetes (Willems, Saunders, Hunt and Schorling (1997)). The main variables of interest are potential risk factors for diabetes, including cholesterol and hypertension, and diagnostic variables, such as glycosolated hemoglobin levels. These variables are believed to be observed as functions of some body configuration measurement like body mass index (weight/height^2) or body weight, which are also measured. Another example is data collected on hemodialysis patients, where albumin and transferrin protein concentrations in plasma are among the variables of main interest (Kaysen et al. (2003)). Similar to the diabetes data, body mass index is again identified as a common confounder. The interest is in obtaining some normalized forms of the protein concentrations that are free from the effect of body mass index and thus are comparable across patients. One simple way of adjustment commonly used in the analysis of such data is through dividing by the confounder, which suggests that the effect of the confounder on the variables is thought to be of a multiplicative nature. Based on this observation, Şentürk and Müller (2005a,b) proposed a more flexible multiplicative adjustment, by modeling the confounding through *unknown functions* of the confounder instead of the confounder itself. This reflects the uncertainty encountered in many applications about the precise nature of the commonly assumed multiplicative relation between the confounder and the variables. For the simple case of two variables of interest, Şentürk and Müller's adjustment models the underlying variables as

$$Y = \frac{\tilde{Y}}{\psi(U)}, \quad X = \frac{\tilde{X}}{\phi(U)}, \tag{1}$$

where they are defined to be the parts of the observed variables, $\tilde{Y}$ and $\tilde{X}$, that are independent of the observable confounder U. Here, $\psi(\cdot)$ and $\phi(\cdot)$ denote unknown smooth contaminating functions of U.

The main goal is to uncover the relationship between the underlying variables of interest, Y and X, based on the observations on the confounder U, and on the contaminated variables, $\tilde{Y}$ and $\tilde{X}$. The need for an adjustment procedure to achieve this goal was demonstrated by Şentürk and Müller (2005a,b). They showed that the multiplicative confounding by U can induce artificial relations between Y and X, which can lead to large biases in estimation. They proposed consistent estimates for the regression coefficients γ_0 and γ_1 in the model

$$Y = \gamma_0 + \gamma_1 X + e \tag{2}$$

(Şentürk and Müller (2005a)), where e is the error term, assumed to be independent of X and U. Consistent estimates for the correlation between Y and X have also been proposed (Şentürk and Müller (2005b)).

In both cases, a common identifiability condition utilized is that the adjustment is mean preserving, i.e. the means of adjusted variables are the same as the means of the observed variables. This identifiability condition amounts to the following constraint on the confounding functions,

$$E\{\psi(U)\} = 1, \qquad E\{\phi(U)\} = 1, \tag{3}$$

under the multiplicative confounding given in (1). It has been shown that under the identifiability condition of no average distortion, the adjustment method proposed for the regression setting, covariate adjusted regression (CAR), is a general adjustment that not only works for multiplicative distortion in (1), but also works for additive (i.e. $Y = \tilde{Y} - \psi(U)$ and $X = \tilde{X} - \phi(U)$) or no distortion (i.e. $Y = \tilde{Y}$ and $X = \tilde{X}$) (Şentürk and Müller (2005a)). Consistent estimates of γ_1 can be obtained for the cases of additive and no distortion by the standard methods of nonparametric partial regression and least squares regression, respectively.

However, there existed no standard procedure that could handle multiplicative confounding, and certainly not one that can handle all three types of confounding together without the need for the specification of the exact form of the confounding. Thus, one of the main attractions of the methodology proposed is that the exact form of the distortion (additive, multiplicative or no distortion) need not be known.

A key observation in formulating the estimation procedure for CAR is that regressing $\tilde{Y}$ on $\tilde{X}$ leads to a varying coefficient model, since both the observed response and predictor vary as a function of the confounder U. This is illustrated in Section 2, where more details on the CAR model are provided. Consistent estimates of the underlying regression coefficients in (2) are then obtained under the identifiability conditions as weighted averages of the estimated coefficients of the above mentioned varying coefficient model. More specifically, the estimation procedure proposed by Şentürk and Müller utilizes equidistant binning (EB) in fitting the varying coefficient model between $\tilde{Y}$ and $\tilde{X}$. In this paper, two new estimation procedures are proposed for CAR, where the main difference to EB is that the equidistant binning is replaced by the nearest neighbor binning (NB) and local polynomial regression (LP) in the proposed algorithms. All three estimation methods are described in detail in Section 3, where the consistency and convergence rates of LP estimators are also given. The proofs are deferred to the Appendix. The simulation study given in Section 4 shows that even though three methods have similar biases, there are substantial differences in their small sample performances, mainly due to their variance. The NB approach improves on EB yielding lower variance for small to moderate sample sizes, and LP proves to be the best estimator among the three with the lowest variance, and therefore mean squared error. The superiority of the local polynomial smoothing compared to the two binning approaches in fitting varying coefficient models affects the overall performance of the estimation of the regression coefficients in (2), leading to estimates with smaller variance. In addition to the lower variance, another advantage of the proposed estimation procedures, NB and LP, is that they are easy to implement with existing software containing least squares procedures. The merging proposed in the implementation of the EB approach is not as standard and straight forward to implement with existing software. We illustrate the proposed estimation method of LP which yields the smallest MSE in the simulation study, with an application to the diabetes data given in Section 5.

2. Covariate Adjusted Regression

Consider the general case of multiple linear regression

$$Y = \gamma_0 + \sum_{r=1}^{p} \gamma_r X_r + e, \tag{4}$$

with p predictors, $X_1, \ldots, X_p$, where e is the error with $E(e) = 0$ and $\mathrm{var}(e) = \sigma^2$, and $\gamma_0, \gamma_1, \ldots \gamma_p$ are the unknown parameters of interest. In the regression model (4), Y and X_r are not observable. Instead, one observes distorted versions $(\tilde{Y}, \tilde{X}_r)$, along with a univariate confounder U, where

$$\tilde{Y}(U) = \psi(U)Y \qquad \text{and} \qquad \tilde{X}_r(U) = \phi_r(U)X_r, \tag{5}$$

for $r = 1, \ldots, p$, and ϕ_r and ψ are unknown smooth functions of U. The identifiability conditions given in (3) can be extended for the case of multiple linear regression as

$$E\{\psi(U)\} = 1, \qquad\qquad E\{\phi_r(U)\} = 1. \tag{6}$$

Model (4) - (6) is the multiplicative distortion or CAR model introduced in Şentürk and Müller (2005a).

A central goal is to obtain consistent estimators of the regression coefficients in model (4), given the observations of the confounding variable U and the distorted observations $(\tilde{Y}, \tilde{X}_r)$ in (5). As previously described in Şentürk and Müller (2005a), the key to the estimation of the targeted regression parameters $\{\gamma_r\}$ is to express the regression of $\tilde{Y}$ on $\{\tilde{X}_r\}_{r=0}^{p}$ as a varying coefficient model. More precisely, under the assumption that $(e,\, U,\, X_r)$ are mutually independent $(r = 1, \ldots, p)$, the regression of $\tilde{Y}$ on $\{\tilde{X}_r\}_{r=0}^{p}$ can be expressed as

$$E(\tilde{Y}|\tilde{X}, U) = \beta_0(U) + \sum_{r=1}^{p} \beta_r(U)\tilde{X}_r,$$

where

$$\beta_0(U) = \psi(U)\gamma_0, \qquad\qquad \beta_r(U) = \gamma_r \frac{\psi(U)}{\phi_r(U)}. \tag{7}$$

Therefore,

$$\tilde{Y} = \beta_0(U) + \sum_{r=1}^{p} \beta_r(U)\tilde{X}_r + \epsilon, \tag{8}$$

with $\epsilon \equiv \psi(U)e$, is a multiple varying-coefficient model (Cleveland, Grosse and Shyu (1991); Hastie and Tibshirani (1993)). Varying coefficient models are an appealing extension of the regression models where the coefficients are allowed to vary as smooth functions of a covariate possibly different than the predictors. They have been popular in diverse application areas, as they reduce the modeling bias with their unique structure while also avoiding the "curse of dimensionality" problem. The literature includes Ramsay and Silverman (1997) on functional data analysis, Nicholls and Quinn (1982) and Chen and Tsay (1993) on nonlinear time series, Wu and Yu (2002) with their overview of applications to longitudinal data. Some approaches to estimation in varying coefficient models for independent and identically distributed data are described in Hoover, Rice, Wu and Yang (1998), Wu and Chiang (2000), Chiang, Rice and Wu (2001), Cai, Fan and Li (2000) and Fan and Zhang (1999).

3. Estimation Procedures

The estimation of the regression coefficients $\{\gamma_r\}_{r=0}^{p}$ in (4) is a two-step estimation procedure. The first step involves estimation of the varying coefficient functions, $\beta_r(\,\cdot\,)$ in (8), which are estimable since $\tilde{Y}$, $\tilde{X}$, and U are all observable. The coefficients $\{\gamma_r\}_{r=0}^{p}$ are targeted in the second step, with weighted averages of the estimated $\beta_r(\,\cdot\,)$, making use of the relations between $\beta_r(\,\cdot\,)$ and γ_r given by (7) and the identifiability conditions. Şentürk and Müller (2005a) use the method of equidistant binning for the estimation of $\beta_r(\,\cdot\,)$ in the first step. They divide the support of U into m bins and then fit linear regressions of $\tilde{Y}$ on $\tilde{X}$ using the data falling within each bin. The regression coefficients estimated in each bin are the raw estimates of $\beta_r(\,\cdot\,)$. In the second step, estimates of $\{\gamma_r\}_{r=0}^p$ are obtained based on weighted averages of these raw estimates coming from each bin. We propose two new estimation procedures where in the first proposed method, the nearest neighbor binning approach is explored in place of equidistant binning. The second proposed method utilizes a more advanced smoothing method, local polynomial modeling in estimating $\beta_r(\,\cdot\,)$ in the first step. The second step of LP involves targeting $\{\gamma_r\}_{r=0}^p$ with weighted averages of the raw estimates of $\beta_r(\,\cdot\,)$ obtained in step 1, this time evaluated at the original observed U_i, $i = 1, \ldots, n$, values.

3.1. Estimation via equidistant binning

Recall that the available (observed) data are of the form $(U_i, \tilde{\mathbf{X}}_i, \tilde{Y}_i)$, $i = 1, \ldots, n$, for a sample of size n, where $\tilde{\mathbf{X}}_i = (\tilde{X}_{1i}, \ldots, \tilde{X}_{pi})^{\mathrm{T}}$ are the p-dimensional predictors. It is assumed that the covariate U is bounded below and above, $a \le U \le b$, where $a < b$ are real numbers. The estimation procedure initially divides the interval $[a, b]$ into m equidistant intervals, $B_1, \ldots, B_m$, referred to as bins. Let L_j be the number of U_i's falling into bin j. Furthermore, denote the data for which $U_i \in B_j$ by the collection $\{(U'_{jk}, \tilde{X}'_{rjk}, \tilde{Y}'_{jk}),\ k = 1, \ldots, L_j, r = 1, \ldots, p\} = \{(U_i, \tilde{X}_{ri}, \tilde{Y}_i), i = 1, \ldots, n, r = 1, \ldots, p : U_i \in B_j\}$, where $(U'_{jk}, \tilde{X}'_{rjk}, \tilde{Y}'_{jk})$ is the kth data element in the jth bin, B_j. Data elements in any given bin are marked by a prime.

After the initial binning of the data, a linear regression is fitted to the data within each bin B_j, $j = 1, \ldots, m$. More precisely, the least squares estimator of the multiple regression of the data in the jth bin is

$$\hat{\boldsymbol{\beta}}_j = (\hat{\beta}_{0j}, \ldots, \hat{\beta}_{pj})^{\mathrm{T}} = (\tilde{\mathbf{X}}_j^{\prime T} \tilde{\mathbf{X}}'_j)^{-1} \tilde{\mathbf{X}}_j^{\prime T} \tilde{\mathbf{Y}}'_j, \tag{9}$$

where $\tilde{\mathbf{X}}'_j = (\tilde{\mathbf{X}}'_{j1}, \ldots, \tilde{\mathbf{X}}'_{jL_j})^{\mathrm{T}}$ is the $L_j \times (p+1)$ data matrix in bin j, with the kth observation $\tilde{\mathbf{X}}'_{jk} = (1, \tilde{X}'_{1jk}, \ldots, \tilde{X}'_{pjk})^{\mathrm{T}}$ and the response vector is $\tilde{\mathbf{Y}}'_j = (\tilde{Y}'_{j1}, \ldots, \tilde{Y}'_{jL_j})^{\mathrm{T}}$.

In the second step of the estimation procedure, the estimators of the targeted regression parameters $\{\gamma_r\}_{r=0}^{p}$ are obtained as weighted averages of the raw estimators $\{\hat{\boldsymbol{\beta}}_j\}_{j=1}^{m}$ from m bins,

$$\hat{\gamma}_{0,EB} = \sum_{j=1}^{m} \frac{L_j}{n} \hat{\beta}_{0j} \qquad \text{and} \qquad \hat{\gamma}_{r,EB} = \frac{1}{\bar{\bar{\tilde{X}}}_r} \sum_{j=1}^{m} \frac{L_j}{n} \hat{\beta}_{rj} \bar{\bar{\tilde{X}}}{}'_{rj}, \tag{10}$$

where $\bar{\bar{\tilde{X}}}_r = n^{-1} \sum_{i=1}^{n} \tilde{X}_{ri}$ and $\bar{\bar{\tilde{X}}}{}'_{rj} = L_j^{-1} \sum_{k=1}^{L_j} \tilde{X}'_{rjk}$ (Şentürk and Müller (2005a)). Note that the weights depend on the number of data points in each bin, namely L_j for $j = 1, \ldots, m$. These estimators are motivated by the relations $E\{\beta_0(U)\} = \gamma_0$ and $E\{\beta_r(U)\tilde{X}_r\} = \gamma_r E\{\psi(U)X_r\} = \gamma_r E(X_r) = \gamma_r E(\tilde{X}_r)$. (These relations follow directly from (6) and (7).)

Bin width or equivalently the total number of bins formed acts as a smoothing parameter for EB. It is suggested that the bin width would be chosen such that the average number of points falling in each bin is $p + 1$, enough to fit the linear regressions, where p is the number of parameters of the regression model. In addition, merging of the sparsely populated bins is introduced for the implementation of the binning algorithm. It entails that if there are bins with less than $p + 1$ elements, such bins would be merged with a neighboring bin. Even though the idea behind merging seems straight forward, its implementation requires careful calibration. In order not to introduce bias into the procedure, issues such as randomization of the order of bins merged and the choice of the neighboring bin that they are merged with have to be considered. This is a potential drawback with the implementation of the EB algorithm.

3.2. Estimation via nearest neighbor binning

As pointed out earlier, for EB, B_j, $j = 1, \ldots, m$, are fixed and equidistant; however, the number of data points, L_j, falling into each bin is random. We explore another binning approach referred to as the nearest neighbor binning. Here, the bin lengths and boundaries are random, but each bin contains the same number of observations, denoted by L. NB utilizes the nearest neighbor idea by first ordering the observed confounder values U_i, $i = 1, \ldots, n$, and then forming the $m = n/L$ number of bins by grouping the sets of L nearest neighbor values among the ordered set starting with the first L to the last. Once the bins are formed, the rest of the procedure is the same as explained for the case of EB. We will denote the estimators for the target regression parameters from NB as $\{\hat{\gamma}_{r,NB}\}_{r=0}^{p}$.

One advantage of NB over EB is that once the number of observations per bin is fixed to $L \geq p+1$, there is no need for merging the bins, because each bin has enough points to fit the linear regressions. This makes the implementation of NB much easier compared to EB.

3.3. Estimation via local polynomial regression

The second proposed approach, LP, estimates the functions $\{\beta_r(U)\}_{r=0}^{p}$ based on local polynomial modeling instead of binning in the first step. (For details on local polynomial modeling, see Fan and Gijbels (1996) and references therein). Fan and Zhang (1999, 2000), Zhang and Lee (2000) and Zhang, Lee and Song (2002) investigated the properties of local polynomial modeling for the estimation in varying coefficient models. The function $\beta_r(U)$ can be approximated based on local polynomial modeling as

$$\beta_r(U) \approx \sum_{k=0}^{q} \frac{1}{k!} \beta_r^{(k)}(u)(U-u)^k, \qquad r = 0, 1, \ldots, p, \tag{11}$$

for U in a neighborhood of u. Here, $\beta_r^{(k)}$ denotes the kth derivative of $\beta_r(\,\cdot\,)$. Consider the local linear least squares estimator of β_r through minimization of

$$\sum_{i=1}^{n} \left[\tilde{Y}_i - \sum_{r=0}^{p} \{\alpha_{r,0} + \alpha_{r,1}(U_i - u)\} \tilde{X}_{ri} \right]^2 K_h(U_i - u), \tag{12}$$

with respect to $\alpha_{r,0}$ and $\alpha_{r,1}$ for a specified kernel function K with bandwidth h where $K_h(\,\cdot\,) = K(\,\cdot/h)/h$. We choose to consider local linear fits for computational simplicity, as they are comparable to local cubic fits for all practical purposes in implementation. Note that $\tilde{X}_{0i} = 1$, corresponding to the intercept function $\beta_0(U)$. Minimization of criterion (12) is a weighted least squares problem. Assuming that $\mathcal{X}^{\mathrm{T}}\mathbf{W}\mathcal{X}$ is nonsingular, the solution is

$$\hat{\boldsymbol{\alpha}} = (\mathcal{X}^{\mathrm{T}}\mathbf{W}\mathcal{X})^{-1}\mathcal{X}^{\mathrm{T}}\mathbf{W}\tilde{\mathbf{Y}},$$

where $\mathcal{X}$ is the following $n \times 2(p+1)$ matrix

$$\mathcal{X} = \begin{bmatrix} 1 & (U_1 - u) & \tilde{X}_{11} & (U_1 - u)\tilde{X}_{11} & \cdots & \tilde{X}_{p1} & (U_1 - u)\tilde{X}_{p1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & (U_n - u) & \tilde{X}_{1n} & (U_n - u)\tilde{X}_{1n} & \cdots & \tilde{X}_{pn} & (U_n - u)\tilde{X}_{pn} \end{bmatrix},$$

$$\mathbf{W} = \mathrm{diag}\{K_h(U_1 - u), \ldots, K_h(U_n - u)\} \quad \text{and} \quad \tilde{\mathbf{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_n)^{\mathrm{T}}.$$

The local least squares estimator of $\beta_r(u)$ is given by

$$\hat{\beta}_r(u) = \mathbf{e}_{2r+1,2(p+1)}^{\mathrm{T}}\hat{\boldsymbol{\alpha}} = \mathbf{e}_{2r+1,2(p+1)}^{\mathrm{T}}(\mathcal{X}^{\mathrm{T}}\mathbf{W}\mathcal{X})^{-1}\mathcal{X}^{\mathrm{T}}\mathbf{W}\tilde{\mathbf{Y}}, \qquad r = 0, \ldots, p, \tag{13}$$

where $\mathbf{e}_{2r+1,2(p+1)}$ is a unit vector of length $2(p+1)$ with 1 in position $2r+1$.

The estimators of the targeted regression parameters, $\{\gamma_r\}_{r=0}^p$, are obtained in the second step by averaging over the raw estimates, $\hat{\beta}_r(U_i)$, this time evaluated at the original observations of the confounder $(U_i)_{i=1}^n$. More precisely,

$$\hat{\beta}_r(U_i) = \mathbf{e}_{2r+1,2(p+1)}^{\mathrm{T}}(\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i\mathcal{X}_i)^{-1}\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i\tilde{\mathbf{Y}}, \tag{14}$$

where $\mathcal{X}_i$ and $\mathbf{W}_i$ are $\mathcal{X}$ and $\mathbf{W}$ with $u = U_i$. This leads to the following regression parameter estimates:

$$\hat{\gamma}_{0,LP} = n^{-1}\sum_{i=1}^n \hat{\beta}_0(U_i) \qquad \text{and} \qquad \hat{\gamma}_{r,LP} = \frac{1}{\bar{\tilde{X}}_r}\sum_{i=1}^n \frac{1}{n}\hat{\beta}_r(U_i)\tilde{X}_{ri},$$

motivated similarly as the binning case, by the relations in (7) and the identifiability conditions in (6). Next, we state the consistency result for the estimators $\{\hat{\gamma}_{r,LP}\}_{r=0}^p$. The proof is given in the Appendix.

Theorem 1. *Under the technical conditions given in the Appendix, it holds that*

$$\hat{\gamma}_{r,LP} = \gamma_r + O_p(n^{-1/2}) + O_p(h^2), \qquad r = 0, \ldots, p.$$

The above result can easily be extended for estimators $\hat{\gamma}_{r,LP}$ obtained based on qth order expansions of the varying coefficient functions $\beta_r(U)$, as given in (11). It would similarly hold that $\hat{\gamma}_{r,LP} = \gamma_r + O_p(n^{-1/2}) + O_p(h^{q+1})$ for $r = 0, \ldots, p$.

The local polynomial modeling approach requires selection of the bandwidth h. For our studies, we utilize the generalized cross-validation (GCV) criterion proposed by Wahba (1977), and Craven and Wahba (1979). Since $\hat{\beta}_r(u)$, given in (13), is linear in $\{Y_i\}_{i=1}^n$, the ith fitted value can be expressed as a linear combination of the response values,

$$\begin{aligned}\hat{\tilde{Y}}_i &= \sum_{j=1}^{n} \{v_j^{(1)}(U_i) + v_j^{(3)}(U_i)\tilde{X}_{1i} + \ldots + v_j^{(2p+1)}(U_i)\tilde{X}_{pi}\}\tilde{Y}_j \\ &\equiv \sum_{i=j}^{n} \{v_j^*(U_i)\}\tilde{Y}_j,\end{aligned}$$

where $\mathbf{v}^{(2r+1)}(U_i) = [v_1^{(2r+1)}(U_i), \ldots, v_n^{(2r+1)}(U_i)] = \mathbf{e}_{2r+1,2(p+1)}^T(\mathcal{X}_i^T\mathbf{W}_i\mathcal{X}_i)^{-1}\mathcal{X}_i^T\mathbf{W}_i$ for a fixed bandwidth h. Furthermore, let $\hat{\tilde{\mathbf{Y}}} = \mathbf{V}\tilde{\mathbf{Y}}$, where the ith row of $\mathbf{V}$ is $[v_1^*(U_i), v_2^*(U_i), \ldots,$ $v_n^*(U_i)]$, $i = 1, \ldots, n$. The bandwidth h is selected to minimize the following GCV criterion, which is a function of the residual sum of squares $RSS(h) = \|\tilde{\mathbf{Y}} - \hat{\tilde{\mathbf{Y}}}\|^2$,

$$h_T = \text{argmin}_h\{T(h)\} = \text{argmin}_h \frac{n^{-1}RSS(h)}{[1 - n^{-1}tr(\mathbf{V})]^2}.$$

Compared to the merging algorithm needed for the implementation of EB, the proposed estimation algorithm via local polynomial regression coupled with the GCV bandwidth choice is much more straight forward to implement. Any standard statistical software package with a least squares routine can be used to obtain $\hat{\beta}_r(U_i)$ in (14) and the estimates $\hat{\gamma}_{r,LP}$, are simply weighted averages of $\hat{\beta}_r(U_i)$.

4. Simulation Study

We compare the finite sample performance of the three estimation methods for CAR via a simulation study. The underlying (unobserved) multiple regression model considered is as follows,

$$Y = 1 + 0.1X_1 + 2X_2 - 0.2X_3 + e, \tag{15}$$

where the predictors X_1, X_2, and X_3 are distributed as $\mathrm{N}(2, 1.5^2)$, $\mathrm{N}(0.5, 0.25^2)$, $\mathrm{N}(1, 1)$, and the error e is distributed as $\mathrm{N}(0, \sigma^2 = 0.25^2)$. For the distribution of the confounding variable U, we consider two cases. For the first case, U is generated from a uniform $[0, 1]$ distribution; and for the second case U is generated from a $\mathrm{N}(6, 1)$ distribution truncated beyond ± three standard deviations. The distorting functions considered are

$$\psi(U) = (U + 3)^2/a, \qquad \phi_1(U) = (U + 10)/b,$$
$$\phi_2(U) = (U + 1)^2/c, \qquad \phi_3(U) = (U + 3)/d,$$

where (a, b, c, d) are $(12.339, 10.5, 2.336, 3.5)$ for $U \sim U[0, 1]$ and $(81.8, 16, 49.8, 9)$ for $U \sim \mathrm{N}(6, 1)$. The constants a, b, c, and d are chosen such that the distorting functions satisfy the identifiability constraints in (6), namely $E\{\psi(U)\} = 1$ and $E\{\phi_r(U)\} = 1$.

For each sample size, n = 50, 70, 100, 150, 200, 400, and 600, we simulate 1000 Monte Carlo data sets. The estimation procedures described in Section 3, EB, NB and LP, are applied to each data set to obtain the estimates $\hat{\gamma}_{r,EB}$, $\hat{\gamma}_{r,NB}$ and $\hat{\gamma}_{r,LP}$ for $r = 0, 1, 2, 3$, respectively. For the estimation procedure based on local polynomial regression, we use a local linear fit ($q = 1$) and the kernel function is taken to be the Epanechnikov kernel, $K(t) = 0.75(1 - t^2)_+$. The bandwidth selection is based on the generalized cross-validation criterion, as previously described in Section 3.3. The range of the bandwidths considered covers up to three standard deviations of the respective confounder distribution. For example, the range considered for the case of a uniform confounder is $0.1, 0.15, \ldots, 0.95$, where the means of the chosen bandwidths are (0.748,0.749,0.670,0.661,0.605,0.427,0.403) corresponding to sample sizes n =(50,70,100,150,200,400,600). For EB, the median number of bins formed after merging are (7,10,14,19,25,39,58) and (7,10,13,18,23,34,51) corresponding to sample sizes

n =(50,70,100,150,200,400,600), for uniformly and normally distributed confounders, respectively. The number of bins used for NB is chosen the be equal to the number of bins formed for EB after merging, for each of the 1000 Monte Carlo runs.

The bias, variance and mean squared error of the estimators have been estimated from the 1000 Monte Carlo runs. In obtaining the estimated bias and variance for the binning method, we excluded four outliers out of 1000 estimates corresponding to a small sample size. The values excluded were four interquartile range away from the median.

All three methods similarly yielded small biases that were negligible compared to their variance for all the sample sizes and confounders considered. More specifically, the squared bias to variance ratios for EB, NB and LP estimators, are all less than 0.5% for the four coefficients and two confounder distributions considered, making the variance the dominating factor in the mean squared error of the estimators. The estimated MSE of the estimators are plotted against the sample size n for the uniformly distributed (left column) and the normally distributed (right column) confounders in Figure 1. The plots for the estimated variance are very similar to those of the estimated MSE, and therefore omitted, since the bias is negligible for this simulation study.

NB yields smaller MSE than EB for both confounders, uniformly for almost all sample sizes. The difference between the two binning procedures is larger for the case of the normal confounder as expected, since the procedures are more similar when the underlying distribution, or the distribution of the confounder is uniform. LP estimators have substantial lower MSE's compared to the two binning algorithms, yielding the best small sample performance among the three estimation procedures considered. For example, in estimation of γ_0, for a small sample size of $n = 50$, the difference in the estimated MSE of LP and EB, and LP and NB are 65% and 63%, of the estimated MSE of EB and of the estimated MSE of NB, respectively. For the moderate sample size of $n = 100$, these percentages are 68% and 58%. As illustrated here, the gain in terms of reduction in MSE by LP can be substantial for small to moderate sample sizes. The MSE's of all three methods get closer to each other as the sample size gets larger (starting from $n = 400$), as expected.

5. Application to Diabetes Data

The diabetes data analyzed here is a subset of a data set that consists of 19 variables collected on 1046 subjects in a study to understand the prevalence of diabetes and other cardiovascular risk factors in Central Virginia for African Americans (Willems, Saunders, Hunt and Schorling (1997)). The data set is available on the web site of the Vanderbilt medical center, department of biostatistics, Vanderbilt university school of medicine. The 215 subjects analyzed here are females who were actually screened for diabetes, where a glycosolated hemoglobin ($GlyHb$) level above 7.0 was taken as a positive diagnosis for diabetes. Interest lies in identifying risk factors for diabetes, among which is hypertension. In this study, body mass index (BMI) was identified to be a factor significantly associated with elevated prevalence of hypertension and diabetes. In order to adjust for this affect of body mass index, the data was analyzed stratified according to BMI, where three groups were formed corresponding to what was considered as low, medium and high BMI values.

We analyze the regression relationship between $GlyHb$ and systolic (SBP) and diastolic (DBP) blood pressures, $GlyHb = \gamma_0 + \gamma_1 SBP + \gamma_2 DBP + e$, adjusted for the confounder BMI directly using the proposed LP estimation for CAR. The CAR estimates are compared to those obtained from the least squares regression of the observed response $\widetilde{GlyHb}$ on the observed predictors $\widetilde{SBP}$ and $\widetilde{DBP}$, i.e. estimation without any adjustment for the confounder U. Four outliers have been removed before analysis, yielding a sample size of $n = 211$. The point estimates and confidence intervals for the regression parameters obtained for both methods are given in Table 1, along with the plots of the estimated coefficient functions from the varying coefficient model $\widetilde{GlyHb} = \beta_0(BMI) + \beta_1(BMI)\widetilde{SBP} + \beta_2(BMI)\widetilde{DBP} + \epsilon(BMI)$ given in Figure 2. In the implementation of CAR with LP, local linear fits ($q = 1$) have been used with Epanechnikov kernel. The bandwidth selection of $h = 10$ is based on generalized cross validation, as described in Section 3.3. The confidence intervals for no adjustment are t confidence intervals for the least squares regression.

The confidence intervals based on CAR with LP are bootstrap percentile confidence intervals given as the $(\alpha/2)B$th and $(1-\alpha/2)B$th percentiles of the bootstrap estimates, $\hat{\gamma}^{(b)}_{0,LP}, \ldots, \hat{\gamma}^{(b)}_{2,LP}$, obtained from $B = 1000$ bootstrap samples generated from the original data. The estimated nonparametric densities of the standardized 1000 bootstrap estimates of γ_0, γ_1 and γ_2 are reasonably close to normal, where they are given in Figure 3, panel 1, overlaying the standard normal density. We also examine the estimated coverage levels of the bootstrap confidence intervals based on LP via simulation. The simulation setting is as described in Section 4 with the normally distributed confounder. For each sample size, n = 50, 70, 100, and 150, 1000 data sets have been generated, where 1000 bootstrap samples have been generated from each data set. Generalized cross validation has been used in bandwidth selection for the bootstrap samples. Figure 3, panel 2, gives the estimated coverage values of the confidence intervals for γ_0 (solid), γ_1 (dash-dotted), γ_2 (dashed) and γ_3 (dotted), corresponding to significance levels of 0.80, 0.90, and 0.95. The estimated coverage values are about 3 to 4 percent above the corresponding levels for the small sample sizes of 50 and 70, but they get closer to the real levels for n =100 and 150.

With the method of no adjustment, DPB is found insignificant at the 0.05 significance level (95% CI: (-0.0476,0.0059)), for $GlyHb$, while SBP is found to be of significant positive predictive value (95% CI: (0.0123, 0.0408)). However, adjusting for BMI with CAR using LP, DBP is found to have a significant negative effect at the 0.05 significance level (95% CI: (-0.0578, -0.0034)), while SBP is found similarly to be of positive predictive value (95% CI:

(0.0135, 0.0405)) for $GlyHb$. The positive effect of SBP increases as BMI values increase from around 20 to 30, and stabilize somewhat for BMI between 30 and 45, with a small dip at around $BMI = 40$ as seen in Figure 2. DBP has a more significant negative effect on $GlyHb$ as the BMI increases, and becomes most significant after $BMI = 38$ for subjects with higher BMI values. While the significant positive predictive effect of SBP on $GlyHb$ found confirms the previous findings that increased body mass index is associated with higher prevalence of hypertension and diabetes (Willems, Saunders, Hunt and Schorling (1997)), the effects of body mass index on both DBP and $GlyHb$ seem to be masking the real negative predictive effect of DBP on $GlyHb$.

6. Discussion

We have proposed two new estimation procedures for CAR, both of which have improved on the earlier proposed EB, in terms of variance, MSE, and ease in implementation. The implementation of NB completely eliminates the need for a merging step, and the implementation of LP is still more straight forward than the merging of EB. The improvement of NB on EB in terms of variance becomes more visible as the confounder distribution deviates from a uniform set-up. Nevertheless, LP, which is also proven to be consistent, improved significantly on the two binning approaches, yielding a much smaller variance. This might be due to the fact that local polynomial modeling is a better smoothing technique in general than binning. In addition, the superior performance of LP shows that in the two-step estimation procedures considered for CAR, the performance of the smoothing technique chosen for estimation of the varying coefficient functions in the first step, does affect the overall performance of the CAR estimates in the second step. With the same argument, it would be of interest to further investigate how some other estimation procedures proposed for varying coefficient models such as smoothing splines, kernel-type estimators, and local maximum likelihood estimates, as previously cited in Section 2, perform when integrated into a suitable estimation scheme for CAR.

Appendix

Technical Conditions

C1. The marginal density $f(U)$ of U has a compact support, say $C(u)$. It has a second continuous derivative and satisfies $\inf_{u \in C(u)} f(u) > 0$ and $\sup_{u \in C(u)} f(u) < \infty$.

C2. The kernel $K(t)$ is a symmetric density function with compact support.

C3. Contamination functions $\psi(\cdot)$ and $\phi_r(\cdot)$, $1 \leq r \leq p$, have continuous second derivatives, $\psi^{(2)}(\cdot) \neq 0$, $\phi_r^{(2)}(\cdot) \neq 0$. They also satisfy $E\psi(U) = 1, \quad E\phi_r(U) = 1, \quad \phi_r(\cdot) > 0$.

C4. For the predictors, $EX_r^{2s} < \infty$ and $E(X_r) \neq 0$; for the errors, $Ee^s < \infty$ for some $s > 2$.

C5. The variables (e, U, X_r) are mutually independent for $r = 1, \ldots, p$.

C6. $h \to 0$, $nh/\log h \to \infty$, and $n^{2\epsilon-1}h \to \infty$ as $n \to \infty$, for some $\epsilon < 1 - s^{-1}$, where s is as given in Condition *C4*.

The following Lemma will be used to prove Theorem 1.

LEMMA 1. *Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent and identically distributed random vectors, where Y_i's are scalar random variables. Assume further that $E|y^s| < \infty$ and $\sup_x \int |y|^s f(x,y)dy < \infty$, where f denotes the joint density of (X, Y). Let K be a bounded positive function with a bounded support, satisfying a Lipschitz condition. Then*

$$\sup_{x \in D} \left| n^{-1} \sum_{i=1}^{n} \{K_h(X_i - x)Y_i - E[K_h(X_i)Y_i]\} \right| = O_p(r_n),$$

provided that $n^{2\epsilon-1}h \to \infty$ for some $\epsilon < 1 - s^{-1}$, where $r_n = [nh/\log(1/h)]^{-1/2}$.

Proof of Lemma 1 follows immediately from the result obtained by Mack and Silverman (1982), as noted by Fan and Zhang (1999).

Proof of Theorem 1

The raw estimates $\hat{\beta}_r(U_i)$ are given by

$$\hat{\beta}_r(U_i) = \mathbf{e}_*^{\mathrm{T}}(\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i\mathcal{X}_i)^{-1}\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i \begin{bmatrix} \sum_{r=0}^{p}\beta_r(U_1)\tilde{X}_{r1}+\epsilon_1 \\ \vdots \\ \sum_{r=0}^{p}\beta_r(U_n)\tilde{X}_{rn}+\epsilon_n \end{bmatrix},$$

where $\mathbf{e}_* \equiv \mathbf{e}_{2r+1,2(p+1)}^{\mathrm{T}}$.

By Taylor's expansion, using condition (C3), we have

$$\begin{aligned}\hat{\beta}_r(U_i) &= \mathbf{e}_*^{\mathrm{T}}(\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i\mathcal{X}_i)^{-1}\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i \begin{bmatrix} \sum_{r=0}^{p}\sum_{k=0}^{1}\beta_r^{(k)}(U_i)(U_1-U_i)^k\tilde{X}_{r1} \\ \vdots \\ \sum_{r=0}^{p}\sum_{k=0}^{1}\beta_r^{(k)}(U_i)(U_n-U_i)^k\tilde{X}_{rn} \end{bmatrix} \\ &+ \mathbf{e}_*^{\mathrm{T}}(\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i\mathcal{X}_i)^{-1}\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i \begin{bmatrix} \sum_{r=0}^{p}\frac{1}{2}\beta_r^{(2)}(\eta_{r1})(U_1-U_i)^2\tilde{X}_{r1}+\epsilon_1 \\ \vdots \\ \sum_{r=0}^{p}\frac{1}{2}\beta_r^{(2)}(\eta_{rn})(U_n-U_i)^2\tilde{X}_{rn}+\epsilon_n \end{bmatrix}\end{aligned}$$

where η_{rj} are between U_i and U_j for $j = 1, \ldots, n$ and $r = 0, \ldots, p$. It follows that

$$\begin{aligned}\hat{\beta}_r(U_i) &= \beta_r(U_i) + \mathbf{e}_*^{\mathrm{T}}(\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i\mathcal{X}_i)^{-1}\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i \begin{bmatrix} \sum_{r=0}^{p}\frac{1}{2}\beta_r^{(2)}(U_i)(U_1-U_i)^2\tilde{X}_{r1} \\ \vdots \\ \sum_{r=0}^{p}\frac{1}{2}\beta_r^{(2)}(U_i)(U_n-U_i)^2\tilde{X}_{rn} \end{bmatrix} \\ &+ \mathbf{e}_*^{\mathrm{T}}(\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i\mathcal{X}_i)^{-1}\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i \begin{bmatrix} \sum_{r=0}^{p}\frac{1}{2}\{\beta_r^{(2)}(\eta_{r1})-\beta_r^{(2)}(U_i)\}(U_1-U_i)^2\tilde{X}_{r1}+\epsilon_1 \\ \vdots \\ \sum_{r=0}^{p}\frac{1}{2}\{\beta_r^{(2)}(\eta_{rn})-\beta_r^{(2)}(U_i)\}(U_n-U_i)^2\tilde{X}_{rn}+\epsilon_n \end{bmatrix} \\ &\equiv \beta_r(U_i) + T_1 + T_2.\end{aligned}$$

Using conditions (C1), (C2), (C6) and the first part of (C4), Lemma 1 can be applied to show that the following holds uniformly in i,

$$\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i\mathcal{X}_i^{\mathrm{T}} = nf(U_i)\mathbf{G}\mathbf{S}_i\mathbf{G}(1+o_p(1))$$

and

$$\mathcal{X}_i^{\mathrm{T}}\mathbf{W}_i \begin{bmatrix} \sum_{r=0}^{p} \frac{1}{2}\beta_r^{(2)}(U_i)(U_1 - U_i)^2 \tilde{X}_{r1} \\ \vdots \\ \sum_{r=0}^{p} \frac{1}{2}\beta_r^{(2)}(U_i)(U_n - U_i)^2 \tilde{X}_{rn} \end{bmatrix} = \frac{n}{2} f(U_i) h^2 \mathbf{G}\boldsymbol{\delta}_i(1 + o_p(1)),$$

where $\mathbf{G} = \mathbf{I}_{p+1} \otimes \mathrm{diag}\{1, h\}$, $\mathbf{S}_i = \mathrm{E}[(\tilde{X}_0, \dots, \tilde{X}_p)^{\mathrm{T}}(\tilde{X}_0, \dots, \tilde{X}_p)|U = U_i] \otimes \mathrm{diag}\{\mu_0, \mu_2\}$, $\boldsymbol{\delta}_i^{\mathrm{T}} = \sum_{r=0}^{p} \beta_r^{(2)}(U_i)[E(\tilde{X}_0\tilde{X}_r|U = U_i), \dots, E(\tilde{X}_p\tilde{X}_r|U = U_i)] \otimes [\mu_2, 0]$ and $\mu_k = \int t^k K(t)dt$. Combining the above two expressions, we have that $T_1 = 2^{-1}h^2\mathbf{e}_*^{\mathrm{T}}\mathbf{S}_i^{-1}\boldsymbol{\delta}_i(1 + o_p(1))$. The term T_2 vanishes by applying Lemma 1, and using the uniform continuity of $\beta_r^{(2)}(\cdot)$, compact support of K, the bounded of the sth moment of the error term e assumed in the second part of (C4), and $E(\epsilon|U) = E(\psi(U)e|U) = 0$. Thus,

$$\hat{\beta}_r(U_i) = \beta_r(U_i) + M_i h^2(1 + o_p(1))$$

uniformly in i, where $M_i = 2^{-1}\mathbf{e}_*^{\mathrm{T}}\mathbf{S}_i^{-1}\boldsymbol{\delta}_i$.

Recalling that $\beta_r(U_i) = \gamma_r\psi(U_i)/\phi_r(U_i)$, the estimator of γ_r based on local polynomial regression is

$$\begin{aligned} \hat{\gamma}_{r,LP} &= \frac{1}{\bar{\tilde{X}}_r}\sum_{i=1}^{n}\frac{1}{n}\hat{\beta}_r(U_i)\tilde{X}_{ri} \\ &= \frac{1}{\bar{\tilde{X}}_r}\sum_{i=1}^{n}\frac{1}{n}\gamma_r\frac{\psi(U_i)}{\phi_r(U_i)}\phi_r(U_i)X_{ri} + \frac{h^2}{\bar{\tilde{X}}_r}\sum_{i=1}^{n}\frac{\tilde{X}_{ri}}{n}M_i(1 + o_p(1)) \\ &= \gamma_r + O_p(n^{-1/2}) + O_p(h^2), \end{aligned}$$

for $r = 1, \dots, p$. This follows from Law of Large Numbers, since $E(\tilde{X}_r) = E(\phi_r(U)X_r) = E(X_r)$, and $E(\psi(U)X_r) = E(X_r)$ follows from (C5) and the identifiability conditions given in (C3). Finally, consistency of $\hat{\gamma}_{0,LP}$ follows from consistency of $\{\hat{\gamma}_{r,LP}\}_{r=1}^{p}$. More precisely,

$$\begin{aligned} \hat{\gamma}_{0,LP} &= n^{-1}\sum_{i=1}^{n}\hat{\beta}_0(U_i) = n^{-1}\sum_{i=1}^{n}\left(\tilde{Y}_i - \hat{\beta}_1(U_i)\tilde{X}_{1i} - \dots - \hat{\beta}_p(U_i)\tilde{X}_{pi}\right) \\ &= \bar{\tilde{Y}} - \hat{\gamma}_1\bar{\tilde{X}}_1 - \dots - \hat{\gamma}_p\hat{\bar{\tilde{X}}}_p = E\tilde{Y} - \sum_{r=1}^{p}\gamma_r EX_r + O_p(n^{-1/2}) + O_p(h^2) \\ &= \gamma_0 + O_p(n^{-1/2}) + O_p(h^2). \end{aligned}$$

References

Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying coefficient models. *Journal of the American Statistical Association* **95**, 888–902.

Chen, R. and Tsay, R. S. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association* **88**, 298–308.

Chiang, C., Rice, J. A. and Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *Journal of the American Statistical Association* **96**, 605–17.

Cleveland, W. S., Grosse, E. and Shyu, W. M. (1991). Local regression models. In *Statistical Models in S* (Chambers, J. M., and Hastie, T. J., eds), 309–376. Wadsworth & Brooks, Pacific Grove.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerical Mathematics* **31**, 377–403.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* London: Chapman and Hall.

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Annals of Statistics* **27**, 1491–1518.

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

Hastie, T. and Tibshirani, R. (1993). Varying coefficient models. *Journal of the Royal Statistical Society, Series B* **55**, 757–96.

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.

Kaysen, G. A., Dubin, J. A., Müller, H. G., Mitch, W. E., Rosales, L. M., Levin, N. W. and the Hemo Study Group (2003). Relationship among inflammation nutrition and physiologic mechanisms establishing albumin levels in hemodialysis patients. *Kidney International* **61**, 2240–2249.

Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **61**, 405–415.

Nicholls, D. F. and Quinn, B. G. (1982). Random coefficient autoregressive models: an introduction. *Lecture Notes in Statistics*, **11**.

Ramsay, J. O. and Silverman, B. W. (1997). *The Analysis of Functional Data.* New York: Springer.

Şentürk, D. and Müller, H. G. (2005a). Covariate adjusted regression. *Biometrika* in press.

Şentürk, D. and Müller, H. G. (2005b). Covariate adjusted correlation analysis. *Scandinavian Journal of Statistics* in press.

Wahba, G. (1977). A survey of some smoothing problems and the method of generalized cross-validation for solving them. In *Applications of Statistics* (ed. P. R. Krisnaiah), pp. 507–523. Amsterdam: North Holland.

Willems, J. P., Saunders, J. T., Hunt, D. E. and Schorling, J. B. (1997). Prevalence of coronary heart disease risk factors among rural blacks: A community-based study. *Southern Medical Journal* **90**, 814–820.

Wu, C. O. and Chiang, C. T. (2000). Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statistica Sinica* **10**, 433–456.

Wu, C. O. and Yu, K. F. (2002). Nonparametric varying coefficient models for the analysis of longitudinal data. *International Statistical Review* **70**, 373–393.

Zhang, W. and Lee, S. Y. (2000). Variable bandwidth selection in varying-coefficient models. *Journal of Multivariate Analysis* **74**, 116–134.

Zhang, W., Lee, S. Y. and Song, X. (2002). Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis* **82**, 166–188.

Table 1: Parameter estimates for the regression model $GlyHb = \gamma_0 + \gamma_1 SBP + \gamma_2 DBP + e$, calculated by least squares regression of $\widetilde{GlyHb}$ on $\widetilde{SBP}$ and $\widetilde{DBP}$ and by covariate-adjusted regression with LP, adjusted for BMI, for 211 subjects. The CIs given are t confidence intervals, and proposed bootstrap CIs, respectively.

	Least Squares Reg.		Covariate-Adjusted Reg.	
Coefficient	Estimate	95% CI	Estimate	95% CI
Intercept	3.6370	(1.7629, 5.5111)	4.5359	(2.6159, 6.3444)
SBP	0.0266	(0.0123, 0.0408)	0.0247	(0.0135, 0.0405)
DBP	-0.0208	(-0.0476,0.0059)	-0.0292	(-0.0578, -0.0034)

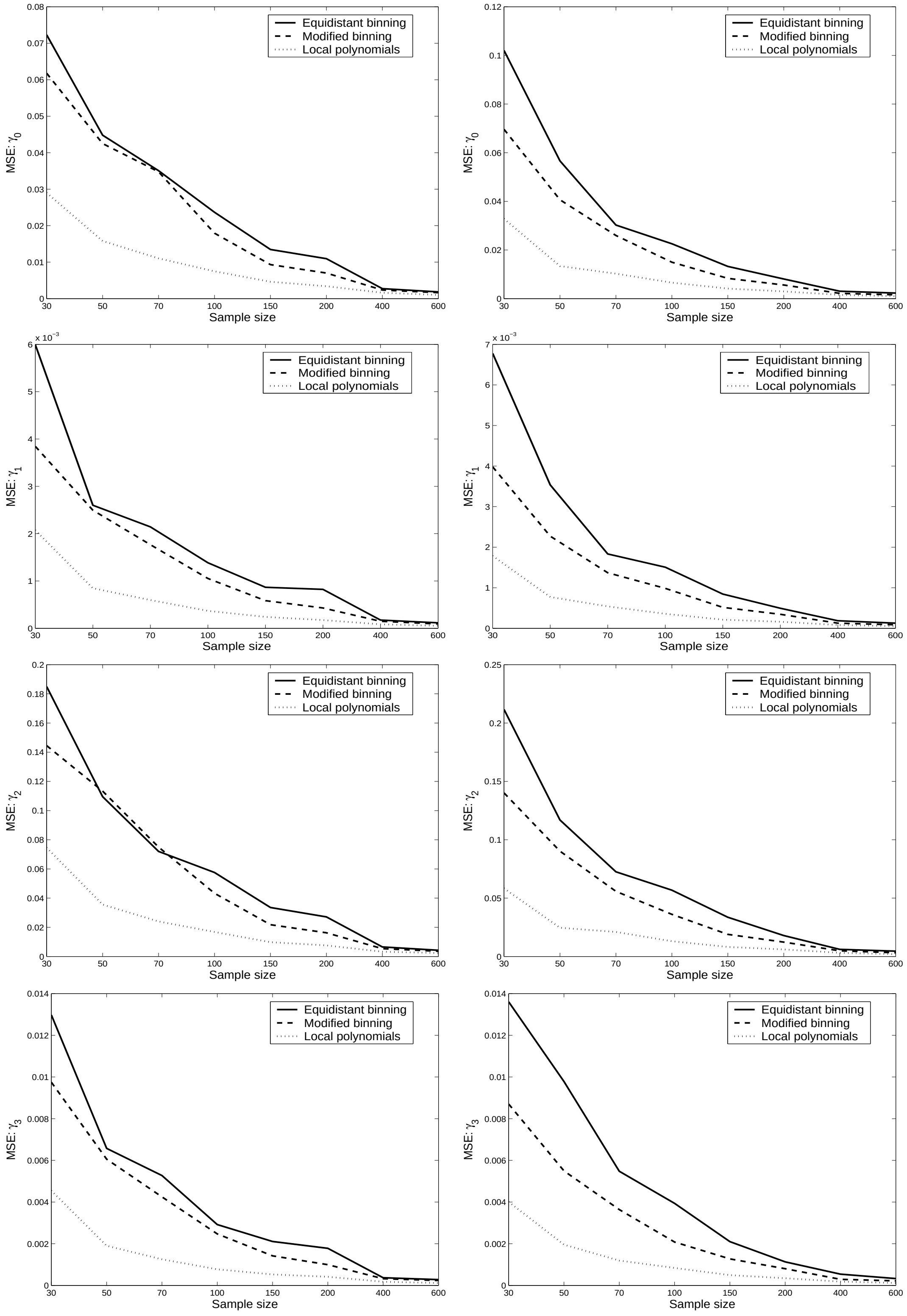

Figure 1: Estimated MSE of the estimators based on two binning algorithms: equidistant binning (NB), modified or nearest neighbor binning (NB), and local polynomial regression(LP) for uniformly distributed (left column) and normally distributed (right column) confounders corresponding to $\gamma_0, \ldots, \gamma_3$ from the simulation model (15).

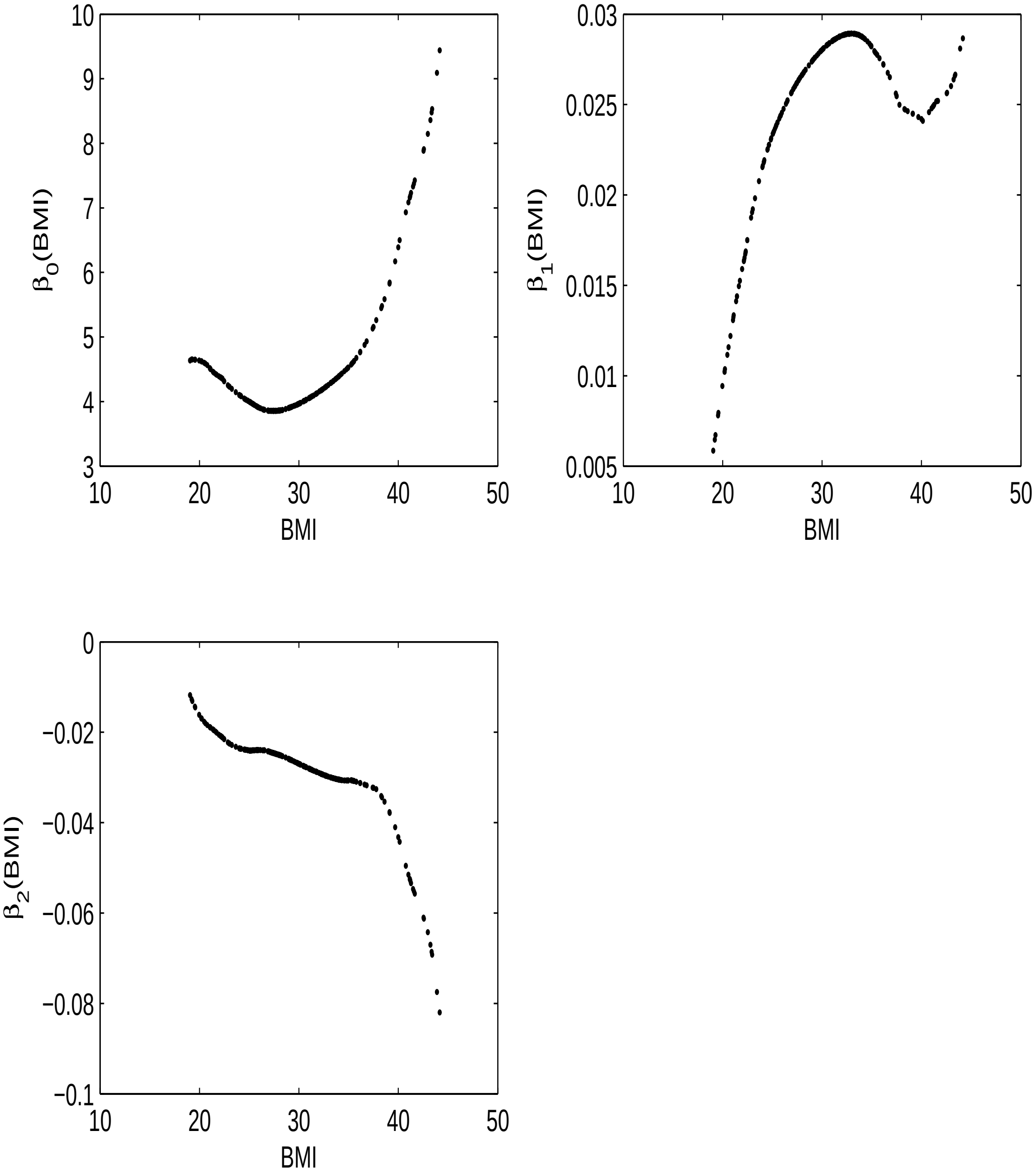

Figure 2: Plots of the estimated smooth coefficient functions $\tilde{\beta}_0(\cdot)$ (top left panel), $\tilde{\beta}_1(\cdot)$ (top right panel) and $\tilde{\beta}_2(\cdot)$ (bottom left panel) for the CAR model $\widetilde{GlyHB} = \beta_0(BMI) + \beta_1(BMI)\widetilde{SBP} + \beta_2(BMI)\widetilde{DBP} + \epsilon(BMI)$ estimated with LP having a generalized cross validation bandwidth choice of $h = 10$. Sample size is 211, and BMI = body mass index, $GlyHb$ = glycosolated hemoglobin level, SBP = systolic blood pressure and DBP = diastolic blood pressure.

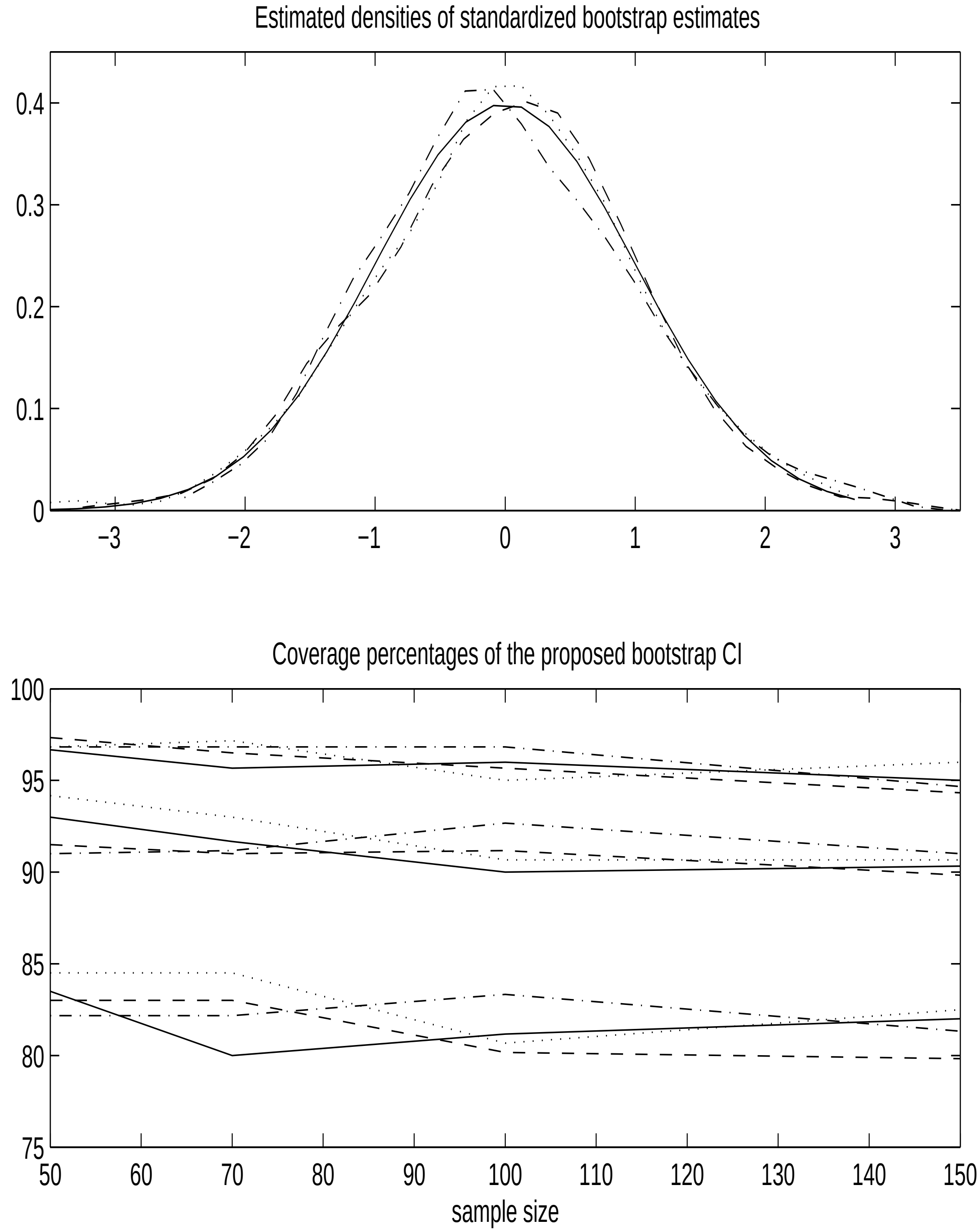

Figure 3: Plot of the estimated nonparametric densities of 1000 bootstrap estimates $\hat{\gamma}^{(b)}_{0,LP}$ (dashed), $\hat{\gamma}^{(b)}_{1,LP}$ (dash-dotted), $\hat{\gamma}^{(b)}_{2,LP}$ (dotted) used in forming 95% CI's of the regression parameters in the analysis of diabetes data, overlaying the standard normal density (solid), (panel 1). A fine binning procedure is followed by local least squares fits with cross validation bandwidth choices of 0.4 to obtain the nonparametric densities. The estimated coverage values of the proposed bootstrap CI's for γ_0 (solid), γ_1 (dash-dotted), γ_2 (dashed), γ_3 (dotted) in the simulation model (15) are given in panel 2, corresponding to significance levels 0.95, 0.90, 0.80.