Published in final edited form as:

J Nonparametr Stat. 2011 December ; 23(4): 917–926. doi:10.1080/10485252.2011.588707.

Double-smoothing for Varying Coefficient Models

Wan Tang^{a,*}, Guoxin Zuo^{a,b}, and Hua He^a

^aDepartment of Biostatistics and Computational Biology, University of Rochester, USA

^bSchool of Mathematics and Statistics, Central China Normal University, China

Abstract

Moderation analyses are widely used in biomedical and psychosocial research to investigate differential treatment effects, with moderators frequently identified through testing the significance of the interaction between the predictor and the potential moderator under strong parametric assumptions. Without imposing any parametric forms on how the moderators may affect the relationship between predictors and responses, varying coefficient models address this fundamental problem of strong parametric assumptions with current practice of moderation analysis and provide a much broader class of models for complex moderation relationships. Local polynomial, especially local linear, methods are commonly used in estimating the varying coefficient models. Recently, a double-smoothing (DS) local linear method has been proposed for nonparametric regression models, with nice properties compared to local linear and local cubic methods. In this paper, we generalize DS to varying coefficient models, and show that it holds similar advantages over local linear and local cubic methods.

Keywords

varying coefficient; double smoothing; bias reduction

1. Introduction

Moderation analyses are widely used in biomedical and psychosocial research to investigate differential treatment effects (Chaplin 1991; Crits-Christoph et al. 1999; Kraemer et al. 2002). In practice, moderators are identified through testing the significance of the interaction between predictors and the potential moderator. However, the validity of this common practice depends on the validity of the model, which assumes that the moderator has a linear effect on the relationship between the predictor and the responses, i.e., the slope is a linear function of the moderator. Obviously this is very restrictive, since in many situations it is unclear how the potential moderator may change the slope, even if the predictor and response do follow a linear relation. In such cases, more complex models that better model the moderation relationship should be applied. Without imposing any parametric assumption, nonparametric smoothing methods provide a curve estimate of the coefficient, which may be used as a guide for (parametric) model selection, and in particular for testing whether the interaction model is appropriate for the moderation relationship (Tang et al. 2009). However, the smoothing coefficients themselves are of interest, and they are studied under varying coefficient models in the field of statistical modeling (Fan and Zhang 1999, 2008; Hastie and Tibshirani 1993; Zhang et al. 2002).

Local polynomial, especially local linear, methods may be the most commonly used nonparametric smoothing technique in nonparametric regression (Eubank 1999; Fan and Gijbels 1996; Ruppert et al. 2003; Wand and Jones 1995). Compared to high-order local polynomial regression, local linear regression is less likely to encounter the sparse data problem because its design matrix is less likely to be singular or nearly singular. In local linear or polynomial regression analysis, the estimates of local constant terms are used as the estimates of fitted values or curves, while other coefficients of local polynomials obtained during estimation, which may be used to estimate derivatives of curves, are completely ignored. Recently, He and Huang (2009) proposed a double-smoothing (DS) method to take into account such information in the estimation of univariate nonparametric regression. For a target point, there is a fitted value for each local line under local linear regression. By integrally combining all these fitted values with another round of smoothing, this DS local linear method can reduce the bias from order h^2 to order h^4 , while keeping the variance at the order $(nh)^{-1}$ (h is the bandwidth or smoothing parameter and n is the sample size) (He and Huang 2009). Thus, DS has better performance than local linear regression in situations where considerable bias exists. The bias and variance of DS are comparable with those of local cubic regression; the orders are the same. However, DS more easily overcomes the sparse data problem, because the design matrices that are used in the first step of smoothing are those for local linear regression. Because of the curse of dimensionality, sparse data is a common problem for applying local cubic regression in practice even with moderate sample sizes. Thus, DS local linear provides a good balance between local linear and cubic regression in addressing these common problems.

Local polynomial regression methods have been generalized to varying coefficient models (Cai et al. 2000; Fan and Zhang 1999, 2008; Hastie and Tibshirani 1993; Zhang and Lee 2000; Zhang et al. 2002). In this paper, we generalize DS local linear methods to multivariate varying coefficient models, with the expectation of similar bias reduction without changing the convergence rate of the asymptotic variance. As in the case of nonparametric smoothing regression, the DS local linear estimate for varying coefficient models has the similar advantage of overcoming the sparse data problem, since the design matrix is only used in the first step of smoothing.

After a brief review of local polynomial estimation in varying coefficient models in Section 2, we introduce the DS local linear methods for varying coefficient models and develop their main properties in Section 3. To illustrate the finite sample performance of the proposed methods and compare them with the ones based on local linear and local cubic alternatives, simulation studies are conducted in Section 4. Finally the paper concludes with a discussion in Section 5.

2. Local Polynomial Estimation in Varying Coefficient Models

Suppose we have independent observations $(\mathbf{X}_i, Y_i, U_i)$ ($1 \leq i \leq n$) from the model:

$$Y_i = \sum_{j=1}^d m_j(U_i) X_{ij} + \varepsilon_i = \mathbf{X}_i^T \mathbf{m}(U_i) + \varepsilon_i \quad (1)$$

where $\mathbf{X}_i = (X_{i1}, \dots, X_{id})^T$ is a vector of d -dimensional independent variables, U_i is the moderation variable, and $\mathbf{m}(U_i) = (m_1(U_i), \dots, m_d(U_i))^T$ are the varying coefficients of $\mathbf{X}_i$. Further, we assume $E(\varepsilon_i | \mathbf{X}_i, U_i) = 0$ and $\text{Var}(\varepsilon_i | \mathbf{X}_i, U_i) = \sigma^2(U_i)$. At a target point u , we want to estimate the varying coefficient $\mathbf{m}(u)$. Local polynomial method with degree p assumes that in a neighborhood of u , $m_j(U_i)$ can be approximated by

$m_j(U_i) \approx \sum_{k=0}^p \frac{1}{k!} m_j^{(k)}(u)(U_i - u)^k \triangleq \sum_{k=0}^p \beta_{jk}(u)(U_i - u)^k$. Then the local polynomial estimate with bandwidth h is obtained by minimizing a weighted sum of squares:

$$\sum_{i=1}^n \left(Y_i - X_{ij} \sum_{j=1}^d \sum_{k=0}^p \beta_{jk}(U_i - u_0)^k \right)^2 K_h(U_i - u), \quad (2)$$

where $K(\cdot)$ is a symmetric density function, i.e., $K \geq 0$, $\int K = 1$, and $K(-u) = K(u)$, and $K_h(u) = K(u/h)/h$. The density function $K(\cdot)$ generally gives more weight to observations closer to u . If the weight function $K(\cdot)$ is supported on a compact interval, say $[-1, 1]$, then it is obvious that only observations in the region $[u - h, u + h]$ will be used to estimate $\mathbf{m}(u)$. Thus the bandwidth h is a smoothing parameter which controls the size of the neighborhood of local smoothing.

Let $(\hat{\beta}_{j0}(u), \hat{\beta}_{j1}(u), \dots, \hat{\beta}_{jp}(u))^T, j = 1, \dots, d$, be the minimizing point of (2). The local polynomial estimate, denoted by $\tilde{\mathbf{m}}(u; p)$, uses only the constant terms $\hat{\beta}_0(u) = (\hat{\beta}_{10}(u), \dots, \hat{\beta}_{d0}(u))^T$ to estimate the value of $\mathbf{m}(\cdot)$ at a point u , i.e., $\tilde{\mathbf{m}}(u; p) = \hat{\beta}_0(u)$. Higher terms may be used to estimate higher derivatives of $\mathbf{m}(u)$; for example, $\hat{\beta}_{j1}(u)$ may be used to estimate the derivative $m'_j(u)$. Asymptotic behaviors of these estimates have been well studied. The biases and variances of general local polynomial estimates are given in (Zhang and Lee 2000). For convenience of later comparison, the asymptotic biases and variances of the local linear and local cubic estimates are presented in the following lemma.

Lemma 2.1: Let $G(u) = E(\mathbf{X}\mathbf{X}^T | U_i = u)$ and $\Omega(u) = E(\mathbf{X}\mathbf{X}^T | U_i = u) f(u)$, where $f(u)$ is the density of the moderator U . For an interior point u , i.e., if $[u - h, u + h]$ is included in the support of the design density,

a. The asymptotic bias and variance of the local linear estimate for (1) are given by

$$\begin{aligned} \text{bias}\{\tilde{\mathbf{m}}(u; 1) | \mathbf{X}_1, \dots, \mathbf{X}_n, U_1, \dots, U_n\} &= h^2 \frac{\mathbf{m}''(u)}{2} \mu_2 + o_p(h^2), \\ \text{var}\{\tilde{\mathbf{m}}(u; 1) | \mathbf{X}_1, \dots, \mathbf{X}_n, U_1, \dots, U_n\} &= (nh)^{-1} v_0 \Omega^{-1}(u) \sigma^2(u) + o_p\{(nh)^{-1}\}, \end{aligned}$$

where $\mu_j = \int u^j K(u) du$ and $v_j = \int u^j K^2(u) du, j = 0, 1, 2, \dots$

b. The asymptotic bias and variance of the local cubic estimate for (1) are given by

$$\begin{aligned} \text{bias}\{\tilde{\mathbf{m}}(u; 3) | \mathbf{X}_1, \dots, \mathbf{X}_n, U_1, \dots, U_n\} &= \frac{h^4}{24} \frac{\mu_2 \mu_6 - \mu_4^2}{\mu_2^2 - \mu_4} \mathbf{m}^{(4)}(u) + o_p(h^4) \\ \text{var}\{\tilde{\mathbf{m}}(u; 3) | \mathbf{X}_1, \dots, \mathbf{X}_n, U_1, \dots, U_n\} &= \frac{1}{nh} \frac{\mu_4^2 v_0 - 2\mu_2 \mu_4 v_2 + \mu_2^2 v_4}{(\mu_2^2 - \mu_4)^2} \Omega^{-1}(u) \sigma^2(u) + o_p\left(\frac{1}{nh}\right). \end{aligned}$$

Since higher order terms of bias for local linear estimates will be needed for the DS local linear estimate to be developed in the next section, we derive the next two terms (h^3 and h^4) in the following lemma.

Lemma 2.2: For an interior point u , the asymptotic bias of the local linear method for (1) is given by

$$\text{bias}\left\{\tilde{\mathbf{m}}(u) | \mathbf{X}_1, \dots, \mathbf{X}_n\right\} = h^2 \frac{\mathbf{m}''(u)}{2} \mu_2 + h^4 \mathbf{a}(u) + o_p(h^4),$$

where

$$\mathbf{a}(u) = \frac{(\mu_2^2 - \mu_4) \mathbf{m}''(u)}{4} (2\Omega^{-1}\Omega'\Omega^{-1}\Omega' - \Omega^{-1}\Omega'')(u) + \frac{1}{24}\mu_4\mathbf{m}^{(4)}(u).$$

The lemma will be proved in the Appendix.

It is worth to note that the main terms in the asymptotic biases for the local linear and local cubic estimates are similar to their counterparts in nonparametric regression. However, the higher terms can be very different. The coefficient of the h^4 terms developed in Lemma 2.2 is much more complicated than that in the nonparametric regression (Fan and Gijbels 1996). The asymptotic variances for both local linear and local cubic estimates change by a factor, $G^{-1}(u)$, compared to the those for nonparametric regression models. However, in the special cases when $d = 1$ and $G(u)$ is a constant, for example if $\mathbf{X}$ and U are independent, then the asymptotic biases including the higher term for h^4 reduce to the same forms as those for nonparametric regression, while the asymptotic variances are different by a constant.

3. Double-Smoothing Estimator and Its Properties

Under the local linear approach, although $\hat{\beta}_1(u)$ may be used to estimate the derivative of $\mathbf{m}(u)$, information from $\hat{\beta}_1(u)$ is generally ignored for estimating $\mathbf{m}(u)$, if interest focuses on the mean function. For univariate nonparametric regression models, He and Huang (2009) developed the DS local linear method which makes use of both $\hat{\beta}_0(u)$ and $\hat{\beta}_1(u)$ in estimating $m(u)$ to achieve bias reduction. We generalize the DS local linear method to multivariate varying coefficient models. Following the idea of He and Huang (2009), we similarly combine all the fitted values of the fitted lines at a target point t and define the general double-smoothing (DS) local linear estimate of $\mathbf{m}(t)$ as:

$$\int \{\hat{\beta}_0(u) + \hat{\beta}_1(u)(t - u)\} L_{h'}(t - u) du,$$

where $\hat{\beta}_0(u)$ and $\hat{\beta}_1(u)$ are obtained from (2) and $L_{h'}(u) = L(u/h')/h'$ with the weight function $L(\cdot)$ satisfying $L(u) \geq 0$ and $\int L(u) du = 1$. The bandwidth h' is a smoothing parameter which controls the size of the neighborhood for the second step of smoothing. For simplicity, in this paper we focus on the case $K = L$ and $h' = h$. Then the DS local linear estimate, denoted by $\hat{\mathbf{m}}(t)$, becomes

$$\hat{\mathbf{m}}(t) = \int \{\hat{\beta}_0(u) + \hat{\beta}_1(u)(t - u)\} K_h(t - u) du. \quad (3)$$

In order to investigate the properties of the DS local linear estimate defined by (3), it is convenient to introduce some conditions.

Conditions (A):

- A1** The design density $f(u)$ has bounded and continuous second derivative.
- A2** The kernel $K(\cdot)$ is a bounded and symmetric probability density function supported on a compact interval, say $[-1, 1]$.
- A3** The mean function $\mathbf{m}(\cdot)$ has bounded and continuous fourth derivative in a neighborhood of u .

A4 $G(u) = E(\mathbf{X}\mathbf{X}^T | U_i = u)$ exists and is positive-definite. Further, it has bounded and continuous fourth derivative in a neighborhood of u .

A5 $\sigma^2(u) = \text{Var}(Y | \mathbf{X} = \mathbf{x}, U = u)$ is a continuous function of u .

Let

$$K_0^*(v, K) = \int K(v-u)K(u)du, \quad K_1^*(v, K) = \int u(v-u)K(v-u)K(u)du \quad \text{and} \quad V(K) = \int \{K_0^*(v) - K_1^*(v)/\mu_2\}^2 dv$$

. Note that $K_0^*(\cdot)$ is the convolution of $K(u)$ with itself, and $K_1^*(\cdot)$ is the convolution of $uK(u)$ and itself. The properties of the asymptotic bias and variance for the DS local linear estimate $\hat{m}(t)$ in (3) are summarized in Theorem 3.1, with their proofs given in the Appendix. Our results hold for an interior point t , i.e., if $[u - 2h, u + 2h]$ is included in the support of the design density. Thus, if the support of the design density is $[a, b]$, then the interior range is $[a + 2h, b - 2h]$. The reason that the interior range is $[a + 2h, b - 2h]$, rather than the conventional $[a + h, b - h]$, is, similar to the explanation in (He and Huang 2009), because for points in $(a + h, a + 2h]$ and $[b - 2h, b - h)$, the DS local linear estimate uses all local linear (LL) estimates at their neighborhoods of radius h , some of which are boundary points for LL.

Theorem 3.1 : *Under conditions (A), and assume $h = h(n) \rightarrow 0$ and $nh \rightarrow \infty$, then the conditional bias and variance at an interior point t are*

$$\begin{aligned} \text{bias}\{\hat{\mathbf{m}}(t) | \mathbf{X}_1, \dots, \mathbf{X}_n, U_1, \dots, U_n\} &= \frac{1}{4}(\mu_2^2 - \mu_4)B(t)h^4 + o_p(h^4), \\ \text{var}\{\hat{\mathbf{m}}(t) | \mathbf{X}_1, \dots, \mathbf{X}_n, U_1, \dots, U_n\} &= \frac{1}{nh}\sigma^2(t)\Omega^{-1}(t)V(K) + o_p\{(nh)^{-1}\}, \end{aligned}$$

where $B(t) = \mathbf{m}^{(4)}(t) + 2\Omega^{-1}\Omega'\mathbf{m}^{(3)}(t) + \Omega^{-1}\Omega''\mathbf{m}^{(2)}(t)$.

Thus, with another round of smoothing over local linear estimates, the DS local linear estimate has greater bias reduction than LL. Similar to that of DS local linear method in smoothing regression, the intuition for bias reduction in Theorem 3.1 is based on the fact that the LL estimate tends to underestimate the regression mean when the curve is concave, and overestimate the mean when the curve is convex.

Nonparametric regressions can be viewed as special cases of varying coefficient models when $d = 1$ and the covariates X_i are constants. Under such point of view, it is easy to check that Theorem 1 in (He and Huang 2009) is a special case of Theorem 3.1 when $d = 1$ and X_i is a constant. More generally, if $d = 1$ and $G(u)$ is a constant, then $B(t)$ has the same expression as that given in Theorem 1 of He and Huang (2009). The condition $G(u)$ is a constant may be approximately true if the association between X_i and U_i is not strong.

Note also that as in the nonparametric regression case, local polynomial estimate can also reduce the asymptotic bias from order h^2 to a higher order as the degree of polynomial increases (Fan and Gijbels 1996). As shown in Theorem 3.1, the DS local linear estimate $\hat{\mathbf{m}}(t)$ has the same order of the asymptotic bias and variance as local cubic (LC) estimate, but similar to LL estimate, it has the advantage over LC estimate of overcoming sparse data since the design matrix constructed in the same way as that for local linear estimate is used only in the first step.

As a trade-off, DS local linear estimate does have more variability compared to LL estimate. However, its variance in general is smaller than that of LC estimate. For example, when $d = 1$ and $G(u)$ is a constant is a constant, the expressions of the bias and variances are the same as those in the nonparametric regression situations. Thus, we may use the expression given in Table 1 of He and Huang (2009) to assess the variabilities of the estimates. Hence, the results presented in Table 2 of He and Huang (2009) in the cases of the Epanechnikov,

Gaussian, and Uniform kernel functions can also be applied to our varying coefficient situation. For the most commonly-used Epanechnikov kernel, the variance of LC estimate increased by a factor of 2.0833 over that of the LL estimate, however, the variance of the DS local linear estimate $\hat{m}(t)$ only increases by a factor of 1.3826, much less than that of LC estimate.

4. Simulation Study

The finite-sample behavior of the DS local linear estimate for varying coefficient models was investigated and comparisons with LL and LC estimates were made via simulation. For the purpose of simplicity and comparison, our simulation studies focus on the case $d = 1$ with simulation setting similar to that studied in (He and Huang 2009). More precisely, we considered the models in Table 1.

The moderator U has a uniform distribution over $[-2, 2]$. In order to assess the impact of the correlation between X and U , we set

$$X = U_0 + \gamma U, \quad (4)$$

where U_0 has a uniform distribution over $[0, 1]$ and is independent with U . The constant γ controls the correlation between X and U ; it is easy to verify that $\text{corr}(X, U) = 2\gamma / \sqrt{4\gamma^2 + 3}$. Specifically, X and U are independent when $\gamma = 0$.

In the simulation study, we compared DS local linear method with LL and LC methods, with the Epanechnikov kernel as the smoothing function. In the simulation, a random sample of size 200 was drawn from each model. To investigate how the three methods perform with different degrees of smoothing, we varied h from 0.08 to 0.50, and calculated the mean integrated square error (MISE) across 1000 Monte Carlo (MC) simulations at different values of h s. The bias, variance, and mean square error (MSE) were computed at 300 grid points evenly distributed on $[-2, 2]$. These points were also used for the trapezoidal numerical integration in the second step of the DS local linear estimate as well as the MISEs for all the three methods. Note that MISEs were computed as the average of MSEs at the grid points over the same region away from boundary for the three methods. In general, the number of grid points should be chosen so that there are enough points within each interval of length the bandwidth h . Because of the numeric integration in DS local linear method, it requires more computer time compared with LL and LC methods. In our simulation study, the computer time used by the DS local linear method was about 8 times that needed for LL method and about 4 times that needed for LC method.

Shown in Figures 1 are plots of the estimated MISEs vs. h corresponding to the varying coefficient models in Table 1 when U and X are independent (γ in (4) equals 0). The results demonstrate that the MISE for the DS local linear estimate is on average lower than that for LL estimate, implying that the DS local linear approach yielded more accurate estimates than the LL method. The bandwidths that minimize the MISE for LL and DS local linear estimates are close and smaller than that for LC estimate. The curves for LC estimates are not as smooth as those for LL and DS local linear estimates, reflecting the fact that the LC method is not as stable because of sparse data issue.

We also investigated the behaviors of the three methods in terms of the actual curve estimates and their variances. In Table 2, the estimated bias and variances of the three methods corresponding to various bandwidth h s are presented. In general, the bias increases and the variance decreases as the bandwidth increases for large samples, and we can find the

trend clearly in Table 2 for larger bandwidths. However, LC estimates using small bandwidths are more likely to encounter the sparse data problems for small and moderate samples and produce very unstable estimates. It is clear that the DS local linear estimate in general has smaller bias but a little bit larger variances, compared with LL estimate. The MISEs of DS local linear estimates are in general smaller than those of LL estimates, confirming our finding in Figure 1. Compared with DS local linear estimate, the LC estimate also produces comparable bias, variance and MISE if optimal bandwidths are used. However, LC method suffers from the sparse data issue as shown in Figure 1.

To assess the impact of the correlation between U and X , we also compared the three methods with different γ s in (4). The results are summarized in the following table

The simulation result shows that higher correlation between U and X are associated with higher variance in the estimates for all the three methods, but the DS local linear method has still the similar advantages over LL and LC methods.

Note that the advantage of bias reduction of DS local linear method over LL method mainly focuses on bumps. In regions where a straight line fits the data well, there is almost no difference among the three approaches. This point is made clear in Figures 2 where the mean estimated curves based on the three methods for Model 3 (optimal bandwidth based on the MISE was used for each estimate) are presented. Similar results were also obtained for the other two models. The DS local linear estimate has elevated bias near the boundary, indicating the need for adjustment of the boundary effect.

5. Discussion

This paper generalized the double-smoothing (DS) local linear method to multivariate varying coefficient models. Like its applications to nonparametric regression, the DS local linear approach for varying coefficient models combines all fitted values of local lines at a target point t with another round of smoothing. By using all local intercepts and slopes, the DS local linear estimate achieves, just like its application in the univariate regression setting, greater bias reduction than LL estimate; it has an asymptotic bias of order h^4 , rather than h^2 as for LL estimate. As a trade-off, the DS local linear estimate has slightly larger asymptotic variance than the LL estimate. However, when compared to LC estimate, DS local linear method has smaller variability, while achieving the same order of asymptotic bias.

The uniformly lower values of MISE over h for the models in our simulation study show that the DS local linear estimate has better overall performance than LL estimate for finite samples. In terms of asymptotic bias and variance, the DS local linear estimate and LC estimate are comparable, but the latter is quite problematic in practice due to the curse of dimensionality and limited sample sizes in real studies. As a result, the DS local linear estimate provides a good balance between LL and LC estimates in this regard.

Some notable limitations of the current paper include the specialty of the varying coefficient models considered (only one moderator used in all the varying coefficients), and the lack of investigation of boundary effect and bandwidth selection. Further study is needed to extend DS local linear to general and semiparametric varying coefficient models as well as to characterize the effects of boundary and bandwidth on the performance of this approach.

Acknowledgments

The study was supported in part by grant 1R21DA027521-01 from the National Institute on Drug Abuse. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse or the National Institutes of Health.

References

- Cai Z, Fan J, Li R. Efficient Estimation and Inferences for Varying-Coefficient Models. *Journal of the American Statistical Association*. 2000; 95:888–902.
- Chaplin W. The next generation of moderator research in personality psychology. *Journal of Personality*. 1991; 59:143–178. [PubMed: 1880698]
- Crits-Christoph P, Siqueland L, Blaine J, Frank A, Luborsky L, Onken L, Muenz L, Thase M, Weiss R, Gastfriend D, et al. Psychosocial treatments for cocaine dependence: National Institute on Drug Abuse collaborative cocaine treatment study. *Archives of General Psychiatry*. 1999; 56:493. [PubMed: 10359461]
- Eubank, RL. *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker; 1999.
- Fan, J.; Gijbels, I. *Local Polynomial Modelling and its Applications*. London: Chapman and Hall; 1996.
- Fan J, Zhang W. Statistical estimation in varying coefficient models. *Annals of Statistics*. 1999; 27:1491–1518.
- Fan J, Zhang W. Statistical methods with varying coefficient models. *Statistics and its Interface*. 2008; 1:179–195. [PubMed: 18978950]
- Hastie T, Tibshirani R. Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1993; 55:757–796.
- He H, Huang LS. Double-smoothing for bias reduction in local linear regression. *Journal of Statistical Planning and Inference*. 2009; 139:1056–1072.
- Kraemer H, Wilson G, Fairburn C, Agras W. Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*. 2002; 59:877. [PubMed: 12365874]
- Ruppert, D.; Wand, MP.; Carroll, RJ. *Semiparametric Regression*. London: Cambridge University Press; 2003.
- Tang W, Yu Q, Crits-Christoph P, Tu XM. A New Analytic Framework for Moderation Analysis — Moving Beyond Analytic Interactions. *Journal of Data Science*. 2009; 7:313–329. [PubMed: 20161453]
- Wand, MP.; Jones, MC. *Kernel smoothing*, Vol. 60 of *Monographs on Statistics and Applied Probability*. London: Chapman and Hall Ltd.; 1995.
- Zhang W, Lee S. Variable bandwidth selection in varying-coefficient models. *Journal of Multivariate Analysis*. 2000; 74:116–134.
- Zhang W, Lee S, Song X. Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis*. 2002; 82:166–188.

Appendix

Technical details

5.1. Proof of Lemma 2.2

By writing the sum of squares (2) in the matrix form $(Y - \mathbf{Z}\beta)^T \mathbf{W} (Y - \mathbf{Z}\beta)$, where $\beta = (\beta_{10}, \dots, \beta_{1p}, \dots, \beta_{d0}, \dots, \beta_{dp})^T$, $Y = (y_1, \dots, y_n)^T$, $\mathbf{W} = \text{diag}(K_h(U_i - u_0))$, and $\mathbf{Z} = (Z_1 \dots Z_d)_{n \times d(p+1)}$ where

$$Z_j = \begin{pmatrix} X_{1j} & (U_1 - u_0) & X_{1j} \cdots & (U_1 - u_0)^p & X_{1j} \\ X_{2j} & (U_2 - u_0) & X_{2j} \cdots & (U_2 - u_0)^p & X_{2j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{nj} & (U_n - u_0) & X_{nj} \cdots & (U_n - u_0)^p & X_{nj} \end{pmatrix}_{n \times (p+1)}, j=1, \dots, d,$$

we obtain the least square estimation of β at u_0 as

$$\widehat{\beta} = (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{W} \mathbf{Y}.$$

It follows that the local polynomial estimation of $(m_1, \dots, m_d)$ at u_0 is

$$\overline{\mathbf{m}}(u_0, p) = (\mathbf{I}_d \otimes \mathbf{e}_{1,p}^T) (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{W} \mathbf{Y},$$

where $\mathbf{e}_{1,p}^T$ is the $(p+1)$ -dimensional unit vector with the first entry equals to 1 and $\otimes$ is the Kronecker product of matrices.

Let

$$\mathbf{G}_p = \text{diag}(1, \dots, h^p), \mathbf{\Omega}^* = \mathbf{\Omega} \mathbf{m}, \mathbf{\Omega}^{*(s)} = \sum_{k=0}^s \binom{s}{k} \mathbf{\Omega}^{(k)} \mathbf{m}^{(k)}, \Gamma_s = (\mu_{s+i+j})_{0 \leq i, j \leq p}, \Lambda_s = (\nu_{s+i+j})_{0 \leq i, j \leq p}, \Gamma_s^* = (\mu_{i+s})_{0 \leq i \leq p}$$

. By simple algebraic computation, we have

$$\mathbf{Z}^T \mathbf{W} \mathbf{Z} = \mathbf{n} \left(\sum_{s=0}^4 \frac{\mathbf{h}^s}{s!} \mathbf{\Omega}^{(s)} \otimes \mathbf{G}_p \Gamma_s \mathbf{G}_p + \mathbf{o}(\mathbf{h}^4) \right), \text{ and } \mathbf{Z}^T \mathbf{W} \mathbf{Y} = \mathbf{n} \left(\sum_{s=0}^4 \frac{\mathbf{h}^s}{s!} \mathbf{\Omega}^{*(s)} \otimes \mathbf{G}_p \Gamma_s^* + \mathbf{o}(\mathbf{h}^4) \right).$$

Therefore the bias of the local linear estimation $\hat{\mathbf{m}}$ at u is

$$\text{bias}(\overline{\mathbf{m}}(u, 1)) = \frac{h^2}{2} \mu_2 \mathbf{m}''(u) + \frac{h^4}{24} [\mu_4 \mathbf{m}^{(4)}(u) + 6(\mu_4 - \mu_2^2) (\mathbf{\Omega}^{-1} \mathbf{\Omega}'' - 2\mathbf{\Omega}^{-1} \mathbf{\Omega}' \mathbf{\Omega}^{-1} \mathbf{\Omega}') \mathbf{m}''(u)] + o_p(h^4).$$

5.2. Proof of Theorem 3.1

Proof of Bias. Bias of DS local linear estimation equals

$$\begin{aligned}
& \text{bias } (\widehat{\mathbf{m}}(t)) \\
&= E(\widehat{\mathbf{m}}(t)) \\
&\quad - \mathbf{m}(t) \\
&= E \left(\int \left(\widehat{\beta}_0(u) \right. \right. \\
&\quad \left. \left. + \widehat{\beta}_1(u)(t - u) \right) K_h(t - u) du \right) - \mathbf{m}(t) \\
&= \int E \left(\left[\widehat{\beta}_0(u) \right. \right. \\
&\quad \left. \left. - \beta_0(u) \right] K_h(t - u) du \right) + \int E \left(\left[\widehat{\beta}_1(u) \right. \right. \\
&\quad \left. \left. - \beta_1(u) \right] (t - u) K_h(t - u) du \right) \\
&\quad - \int \left[\beta_2(u) (t - u)^2 \right. \\
&\quad \left. + \beta_3(u) (t - u)^3 \right. \\
&\quad \left. + \beta_4(u) (t - u)^4 \right] K_h(t - u) du + o_p(h^4).
\end{aligned}$$

The first term is

$$\frac{h^2}{2} \mu_2 \mathbf{m}''(t) + \frac{h^4}{24} \left[\mu_4 \mathbf{m}^{(4)}(t) + 6(\mu_4 - \mu_2^2) (\Omega^{-1} \Omega'' - 2\Omega^{-1} \Omega' \Omega^{-1} \Omega') \mathbf{m}''(t) \right] + \frac{h^4}{4} \mu_2^2 \mathbf{m}^{(4)}(t) + o_p(h^4),$$

The second term equals to

$$-\frac{h^4}{6} \left[\mu_4 \mathbf{m}^{(4)}(t) + 3(\mu_4 - \mu_2^2) \left((\Omega^{-1} \Omega')' \mathbf{m}''(t) + \Omega^{-1} \Omega' \mathbf{m}^{(3)}(t) \right) \right] + o_p(h^4),$$

The last term is

$$-\frac{h^2}{2} \mu_2 \mathbf{m}''(t) - \frac{h^4}{4} \mu_4 \mathbf{m}^{(4)}(t) + \frac{h^4}{6} \mu_4 \mathbf{m}^{(4)}(t) - \frac{h^4}{24} \mu_4 \mathbf{m}^{(4)}(t) + o_p(h^4).$$

Therefore,

$$\text{bias } (\widehat{\mathbf{m}}(t)) = \frac{h^4}{4} (\mu_4 - \mu_2^2) \left[\mathbf{m}^{(4)}(t) + 2(\Omega^{-1} \Omega') \mathbf{m}^{(3)}(t) + (\Omega^{-1} \Omega'') \mathbf{m}^{(2)}(t) \right],$$

Proof of Variance. With changes of variables $w = (t - u)/h$ and $v_i = (t - U_i)/h$, we have

$$\widehat{\mathbf{m}}(t) = \sum_{i=1}^n \int [I_d \otimes (1, t-u)] (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1} \left((X_{1i}, \dots, X_{di})^T \otimes (1, U_i - u)^T \right) K_h(U_i - u) K_h(t - u) du Y_i = \frac{1}{nh} \sum_{i=1}^n \Omega^{-1}(t) (X_{1i}, \dots, X_{di})^T (K_0^*(v_i) - K_1^*(v_i)/\mu_2) Y_i.$$

Because of the independence between Y_i and Y_j for $i \neq j$, we have

$$\begin{aligned} \text{var}(\widehat{\mathbf{m}}(t)) &= \frac{1}{(nh)^2} \Omega^{-1}(t) \sum_{i=1}^n (X_{1i}, \dots, X_{di})^T (X_{1i}, \dots, X_{di}) \Omega^{-1}(t) \times (K_0^*(v_i) - K_1^*(v_i)/\mu_2)^2 \text{var}(Y_i) \\ &= (nh)^{-1} \Omega^{-1}(t) V(K) \sigma^2 + o_p((nh)^{-1}). \end{aligned}$$

Figure 1. MISEs across 1000 realizations for the DS, LL, and LC estimates.

Figure 2.
The true and average fitted curves based on 1000 Monte Carlo simulations for Model 3.

Table 1

Simulation models

Model	$m(u)$	σ	$\approx$ noise/signal
1	$u + 2e^{-16u^2}$	0.3	0.40
2	$\sin(2u) + 2e^{-16u^2}$	0.1	0.20
3	$0.3e^{-4(u+1)^2} + 0.7e^{-16(u-1)^2}$	0.1	0.643

Table 2

The estimated bias, variances, and MISEs for different h for Model 2.

h	0.08	0.14	0.20	0.26	0.32	0.38	0.44	0.50
MISE	LL	0.0195	0.0052	0.0065	0.0128	0.0247	0.0443	0.0710
	DS	0.0148	0.0053	0.0035	0.0044	0.0087	0.0184	0.0338
	LC	0.1317	0.0181	0.0064	0.0041	0.0035	0.0044	0.0071
Bias	LL	0.0119	0.0299	0.0613	0.1016	0.1484	0.2026	0.2589
	DS	0.0073	0.0055	0.0185	0.0435	0.0789	0.1242	0.1731
	LC	0.0115	0.0047	0.0040	0.0107	0.0224	0.0422	0.0677
Var	LL	0.0194	0.0043	0.0027	0.0024	0.0027	0.0032	0.0039
	DS	0.0148	0.0053	0.0031	0.0025	0.0025	0.0030	0.0038
	LC	0.1316	0.0181	0.0064	0.0040	0.0030	0.0026	0.0025

Table 3

The estimated bias, variances, and MISEs of LL, LC, and DS at optimal h s for different association between U and X (Model 1).

β	Bias			Variance			MISE		
	LL	DS	LC	LL	DS	LC	LL	DS	LC
0	0.06528	0.05605	0.05281	0.01940	0.01801	0.01978	0.02366	0.02115	0.02257
.25	0.08040	0.06611	0.06137	0.02329	0.02162	0.02225	0.02976	0.02599	0.02602
.75	0.06340	0.07000	0.04785	0.02423	0.02141	0.02834	0.02825	0.02631	0.03063
2	0.08381	0.06236	0.06449	0.03022	0.03354	0.04439	0.03725	0.03743	0.04855