Published in final edited form as:

J Nonparametr Stat. 2015 April ; 27(2): 195–213. doi:10.1080/10485252.2015.1026903.

Parametrically guided estimation in nonparametric varying coefficient models with quasi-likelihood

Clemontina A. Davenport, Arnab Maity, and Yichao Wu*

Department of Statistics, North Carolina State University, Raleigh, NC 27695, U.S.A

Abstract

Varying coefficient models allow us to generalize standard linear regression models to incorporate complex covariate effects by modeling the regression coefficients as functions of another covariate. For nonparametric varying coefficients, we can borrow the idea of parametrically guided estimation to improve asymptotic bias. In this paper, we develop a guided estimation procedure for the nonparametric varying coefficient models. Asymptotic properties are established for the guided estimators and a method of bandwidth selection via bias-variance tradeoff is proposed. We compare the performance of the guided estimator with that of the unguided estimator via both simulation and real data examples.

Keywords

generalized linear models; local polynomial smoothing; nonparametric regression; parametrically guided estimation; varying coefficient model

1. Introduction

A common scenario in studies is when the researcher observes some covariates and a response and wants to estimate the conditional mean function of the response given the covariates. A typical approach is to fit a generalized linear model (GLM) (Nelder and Wedderburn 1972) where a parametric assumption is made on a transformation of the conditional mean. This approach is easily interpreted and efficient if the correct parametric model is chosen, but can have serious consequences when the model is misspecified, a common problem in real-world applications. Alternatively, nonparametric methods make little or no assumptions of the model and are robust to model misspecification, however, they are slower to converge (Glad 1998) and can fail when the dimensions of the covariates are too high (Fan and Zhang 2008). In this paper, we combine the benefits of both parametric and nonparametric methods by considering estimation of nonparametric varying coefficient models using a pre-specified parametric family of functions as a guide. We use a local likelihood kernel smoothing based estimation procedure and investigate the asymptotic properties of the resulting estimates. We evaluate the finite sample performance of our method via a simulation study and demonstrate it by applying to the AIDS Clinical Trials Group (ACTG) Protocol 315 data described in Section 5.

*Corresponding author. wu@stat.ncsu.edu.

Varying coefficient models (VCMs) are nonparametric generalized additive models (Wood 2006) that increase the flexibility of linear models by using a smooth function to model the parameters. There are numerous advantages of using VCMs over their parametric linear model counterparts. First, fitting a standard linear model is too restrictive because few real world problems satisfy the assumption of linearity. The true, possibly nonlinear, relationship between the covariates are not well captured by polynomial fitting which can lead to large bias in estimation. Second, VCMs allow for the interaction between covariates to be modeled in a nonparametric way. Models with only main effects included may miss the effect from the interaction between covariates. Another advantage for using VCMs is that the dimensionality issue is avoided because the coefficient functions are allowed to vary as a function of another covariate (Hastie and Tibshirani 1993). VCMs are easy to interpret and arise when it is desirable to know how regression coefficients change over groups, or over time in longitudinal studies. These flexible models can be applied to a variety of data types, including longitudinal data (Brumback and Rice 1998; Hoover, Rice, Wu, and Yang 1998), time series data (Chen and Tsay 1993; Huang and Shen 2004), environmental data (Fan and Zhang 1999), and genetic data (Ma, Yang, Romero, and Cui 2011).

VCMs consist of smooth, functional parameters that need to be estimated and this can be done using penalized spline approaches (Eilers and Marx 1996; Cao, Lin, Wu, and Yu 2010), basis expansion methods (Holdeman 1969), or by applying regression locally (Chen and Tsay 1993; Cai, Fan, and Li 2000). In this paper, we utilize the latter method since these estimators are efficient and have nice sampling properties (Fan 1993). This method involves using a kernel function to weight the likelihood and applying polynomial regression locally using Taylor series approximations. In practice, the full likelihood may be unknown or difficult to construct, so we replace it with the quasi-likelihood introduced by Wedderburn (1974). With the quasi-likelihood, only the relationship between the mean and the variance needs to be specified, and the model will still retain most of the efficiency of a maximum likelihood estimation procedure.

Nonparametric estimators can be enhanced by using a parametric guide. In practice, previous information or exploratory analysis may give some insight on the shape of the unknown functions and this information can be used to speed up convergence and reduce bias. The basic process is as follows: 1) identify a parametric family of functions that captures the shape, 2) remove the trend and carry out local polynomial estimation, and 3) add the trend back to obtain the final estimators of the functions. This guided estimation scheme was first studied in the density estimation framework where Hjort and Glad (1995) showed that their estimator had better bias and similar variance compared to the traditional nonparametric estimator, even when the guide was superficial. It has also been studied in least squares regression (Glad 1998; Martins-Filho, Mishra, and Ullah, 2008) quasi-likelihood models (Fan, Wu, and Feng 2009) and nonparametric additive models (Fan, Maity, Wang, and Wu 2013).

In this paper, we propose a way to improve the coefficient functions in VCMs by using guided techniques. Cai et al. (2000) first proposed local estimation techniques for VCM coefficient functions using local-likelihood equations, and present asymptotic properties of their estimators. Fan et al. (2009) propose using quasi-likelihood models as a general case

when the likelihood is unavailable and present guided estimation when a single covariate and response are observed. Fan et al. (2013) extend this methodology to an additive model where multiple covariates are observed but do not interact with each other. The main contribution of this paper is that we allow for interaction between two observed covariates and thus extend previous works to VCMs, of which Fan et al. (2009) is a special case. We borrow the generality of the quasi-likelihood models and the idea of parametrically guided estimation to improve the bias of our estimators, but we extend the optimal bandwidth selection methodology for the kernel weight function when we apply local polynomial regression. We also give asymptotic theory and results for the guided estimators in the VCM framework, different than the GLM and the additive model in Fan et al. (2009) and Fan et al. (2013), respectively. We show in simulations that the guided estimators have lower bias and similar variance when a fixed bandwidth is used, and lower bias and variance when the optimal bandwidth is used. We then estimate the functional parameters in the (ACTG) Protocol 315 data.

The rest of this paper is organized as follows: in Section 2 we give an overview of QLMs and the standard nonparametric estimation procedure using local polynomial fitting. We propose our parametrically guided estimation scheme in Section 2.2 using two different types of corrections, and give some asymptotic properties of our estimators. In Section 3, we present one method of choosing the bandwidth parameter in local polynomial fitting. We evaluate the performance of our estimators compared to the standard ones in Section 4 and found that our estimators had lower bias when a fixed bandwidth was used, and lower bias and variability when the optimal bandwidth was chosen. We then applied our methodology to the ACTG data in Section 5 and provide some concluding remarks in Section 6.

2. Guided Estimation for Varying Coefficient Models

Assume that for each of n subjects we observe covariates $\mathbf{X}_i = (1, X_{i1}, \dots, X_{iq})^T$ and T_i , and a response Y_i . The VCM for these covariates is defined as

$$\begin{aligned} g(\mu_i) &= \theta_0(T_i) + X_{i1}\theta_1(T_i) + \dots + X_{iq}\theta_q(T_i) \\ &= \mathbf{X}_i^T \boldsymbol{\theta}(T_i), \end{aligned} \quad (1)$$

where $\mu_i = E(Y_i | \mathbf{X}_i, T_i)$ is the conditional mean of the response, $g(\cdot)$ is a link function from the GLM framework, and $\boldsymbol{\theta}(\cdot) = \{\theta_0(\cdot), \theta_1(\cdot), \dots, \theta_q(\cdot)\}^T$ are unknown, smooth functions. The first term models the unique effect of T and the remaining terms model the interaction between $\mathbf{X}$ and T . This VCM is more flexible than a linear regression model because it allows the effect of $\mathbf{X}$ to vary smoothly with T and the effect of T is not restricted to a linear assumption. The goal is to estimate $\boldsymbol{\theta}(\cdot)$ and there are several ways to do this (Cleveland, Grosse, and Shyu 1991; Hastie and Tibshirani 1993). The method we adopt is to use local-likelihood kernel smoothing using a quasi-likelihood.

2.1. Framework and Local Likelihood Estimation

Quasi-likelihood models (QLMs), an extension of GLMs, are ideal because often the full likelihood may be unknown or difficult to construct. In QLMs, only the relationship between

the conditional mean and variance of the responses need to be specified, which is often doable in practice. The full conditional log-likelihood is replaced with a quasi-likelihood function $Q(Y_i, \mu_i)$ and if we define $\text{var}(Y_i/\mathbf{X}_i, T_i) = V(\mu_i)$, then Q satisfies

$$\frac{\partial}{\partial \mu} Q(y, \mu) = \frac{y - \mu}{V(\mu)}.$$

Wedderburn (1974) shows that Q has similar properties to the log-likelihoods and that Q is exactly the likelihood when the response comes from a single parameter exponential family.

The standard, unguided, nonparametric procedure for estimating $\theta(\cdot)$ is to use local polynomial fitting by first approximating the functions using a Taylor series expansion so that

$$\theta_j(T_i) \approx \beta_{j0} + (T_i - t_0)\beta_{j1} + \cdots + (T_i - t_0)^P \beta_{jP}, \quad (2)$$

where $\beta_{jp} = \theta_j^{(p)}(t_0)/p!$ for $p = 0, \dots, P$ and $j = 0, \dots, q$. Substituting this approximation into

(1) yields $g(\mu_i) = \mathbf{G}_i^T \boldsymbol{\beta}$ where $\mathbf{G}_i = \{1, (T_i - t_0)^1, \dots, (T_i - t_0)^P\}^T \otimes \mathbf{X}_i$, and

$\boldsymbol{\beta} = (\boldsymbol{\beta}_0^T, \dots, \boldsymbol{\beta}_p^T)^T$ where $\boldsymbol{\beta}_p = (\beta_{0p}, \dots, \beta_{qp})^T$. The approximation in (2) is only accurate when t_0 is close to T_i , so the quasi-likelihood is weighted in such a way that more weight is given to t_0 's close to T_i and little to no weight to those far from T_i . This is done by using a kernel function $K_h(\cdot) = K(\cdot/h)/h$ and defining the local quasi-likelihood as

$$\sum_{i=1}^n Q\{g^{-1}(\mathbf{G}_i^T \boldsymbol{\beta}), Y_i\} K_h(T_i - t_0). \quad (3)$$

The parameter h is the bandwidth and needs to be estimated (see Section 3). We maximize (3) with respect to $\boldsymbol{\beta}$ and the solution $\hat{\boldsymbol{\beta}}_0$ will be the estimate of $\theta(t_0)$.

2.2. Guided Estimation

The local likelihood estimators can be enhanced by using a parametric guide. Intuitively speaking, more curvature in a function makes it more difficult to estimate. If, through exploratory analysis or prior information, one has some idea of the shape of the true function, then one can identify a parametric family that captures this trend. Using the parametric guide, the curvature of the function can be removed yielding a flatter curve that is easier to estimate. Once this flatter curve is estimated, then the guide can be used to add the trend back and obtain the final estimate of the original function. This type of guided estimation has been shown to reduce bias of nonparametric estimators (Hjort and Glad 1995; Glad 1998; Martins-Filho et al. 2008) and improve variance since a larger bandwidth can be selected (Fan et al. 2009, 2013).

Define a parametric family that captures the trend of the function as $\{\theta_{jg}(t, \alpha_j): \alpha_j = (\alpha_{j1}, \dots, \alpha_{j m_j})^T \in \mathbb{A} \subset \mathbb{R}^{m_j}\}$ for $j = 0, \dots, q$. The optimal guides can be found by maximizing the quasi-likelihood

$$\sum_{i=1}^n Q[g^{-1}\{\mathbf{X}_i^T \boldsymbol{\theta}_g(T_i, \boldsymbol{\alpha})\}, Y_i]$$

with respect to $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_0^T, \dots, \boldsymbol{\alpha}_q^T)^T$ where $\boldsymbol{\theta}_g(T_i, \boldsymbol{\alpha}) = \{\theta_{0g}(T_i, \boldsymbol{\alpha}_0), \dots, \theta_{qg}(T_i, \boldsymbol{\alpha}_q)\}^T$. The best fit is denoted by $\hat{\boldsymbol{\theta}}_g(t) = \boldsymbol{\theta}_g(t, \hat{\boldsymbol{\alpha}})$ where we suppress the dependence on $\boldsymbol{\alpha}$ in our notation. In this paper we present two methods of removing the trend using an additive correction or a multiplicative correction.

2.2.1. Additive Correction—If the curvature of θ_j is well approximated by $\hat{\theta}_{jg}$, then estimating the quantity $\theta_j(t) - \hat{\theta}_{jg}(t)$ will yield more accurate and less variable estimates since this quantity is close to flat. Once estimated, the guide is added back to give the final estimate of θ_j . This process can be achieved in one step by defining $\eta_j(t) = \theta_j(t) - \hat{\theta}_{jg}(t) + \hat{\theta}_{jg}(t_0)$ and estimating η_j at t_0 . This definition of η_j is known as the additive correction.

Using this correction, the VCM in (1) can be rewritten as

$$g(\mu_i) = \mathbf{X}_i^T \boldsymbol{\eta}(T_i) + \mathbf{X}_i^T \mathbf{h}(T_i)$$

where $\boldsymbol{\eta}(t) = \{\eta_0(t), \dots, \eta_q(t)\}^T$, and $\mathbf{h}(t) = \{\hat{\theta}_{0g}(t) - \hat{\theta}_{0g}(t_0), \dots, \hat{\theta}_{qg}(t) - \hat{\theta}_{qg}(t_0)\}^T$. Similar to Section 2.1 and with a slight abuse of notation, a Taylor series approximation is used for $\eta_j(T_i)$ about the point t_0 such that

$$\eta_j(T_i) \approx \beta_{j0} + (T_i - t_0)\beta_{j1} + \dots + (T_i - t_0)^P \beta_{jP}.$$

where $\beta_{jp} = \eta_j^{(p)}(t_0)/p!$ for $p = 0, \dots, P$. The local quasi-likelihood

$$\begin{aligned} Q(\boldsymbol{\beta}) &\equiv Q(\boldsymbol{\beta}; h, t_0, \hat{\boldsymbol{\alpha}}) \\ &= \sum_{i=1}^n Q[g^{-1}\{\mathbf{G}_i^T \boldsymbol{\beta} + \mathbf{X}_i^T \mathbf{h}(T_i)\}, Y_i] \times K_h(T_i - t_0) \end{aligned}$$

is maximized with respect to $\boldsymbol{\beta}$ and the estimate of $\hat{\boldsymbol{\beta}}_0$ corresponds to $\hat{\boldsymbol{\eta}}(t_0)$ which gives the final estimate $\hat{\boldsymbol{\theta}}(t_0)$. Because $\mathbf{h}(t)$ is known for fixed $T = t$, our model can be fit using standard software with $\mathbf{h}(t)$ as an offset.

2.2.2. Multiplicative Correction—An alternative correction which leads to a different guided estimator is the multiplicative correction. As in the additive case, the ratio $\theta_j(t)/\hat{\theta}_{jg}(t)$ will be flat if $\hat{\theta}_{jg}(t)$ captures the trend of $\theta_j(t)$ and estimating this ratio will be less biased

than estimating the unknown function directly. Once estimated, the ratio is then multiplied by the guide to get the final estimate of θ_j . The multiplicative correction is defined as $\eta_j(t) = \theta_j(t) \{ \hat{\theta}_{jg}(t_0) / \hat{\theta}_{jg}(t) \}$ and the one step solution requires estimating $\eta_j(t)$ at t_0 .

An alternative correction which leads to a different guided estimator is the multiplicative correction. As in the additive case, the ratio $\theta_j(t) / \hat{\theta}_{jg}(t)$ will be flat if $\hat{\theta}_{jg}(t)$ captures the trend of $\theta_j(t)$ and estimating this ratio will be less biased than estimating the unknown function directly. Once estimated, the ratio is then multiplied by the guide to get the final estimate of θ_j . The multiplicative correction is defined as $\eta_j(t) = \theta_j(t) \{ \hat{\theta}_{jg}(t_0) / \hat{\theta}_{jg}(t) \}$ and the one step solution requires estimating $\eta_j(t)$ at t_0 .

Using the multiplicative correction, (1) is written as

$$g(\mu_i) = \frac{\hat{\theta}_{0g}(T_i)}{\hat{\theta}_{0g}(t_0)} \eta_0(T_i) + \frac{\hat{\theta}_{1g}(T_i)}{\hat{\theta}_{1g}(t_0)} X_{i1} \eta_1(T_i) + \cdots + \frac{\hat{\theta}_{qg}(T_i)}{\hat{\theta}_{qg}(t_0)} X_{iq} \eta_q(T_i).$$

Estimating $\eta_j(T)$ is achieved by first using a Taylor series expansion of $\eta_j(t)$ about the point t_0 and then maximizing

$$Q(\beta) = \sum_{i=1}^n Q \{ g^{-1}(\mathbf{G}_i^* \beta), Y_i \} K_h(T_i - t_0)$$

with respect to β , where

$\mathbf{G}_i^* = \{1, T_i - t_0, \dots, (T_i - t_0)^p\} \otimes \left(\frac{\hat{\theta}_{0g}(T_i)}{\hat{\theta}_{0g}(t_0)}, X_{i1} \frac{\hat{\theta}_{1g}(T_i)}{\hat{\theta}_{1g}(t_0)}, \dots, X_{iq} \frac{\hat{\theta}_{qg}(T_i)}{\hat{\theta}_{qg}(t_0)} \right)$, and $\otimes$ denotes the Kronecker product between the two vectors. The solution $\hat{\beta}_0$ give our final estimates of $\hat{\theta}(t_0)$. By manipulating the design matrix, there is no offset for the multiplicative correction and this model can easily be fit using standard GLM software.

2.3. Asymptotic Properties

In this section, we investigate the asymptotic properties of the proposed guided estimators. For illustration purposes, we present the case where estimation is performed using a local linear approximation ($P = 1$) and an additive guide. Similar results for general P and the multiplicative guide can be obtained by straightforward but tedious algebra.

Define $\kappa_d = \int u^d K(u) du$ and $\nu_d = \int u^d K^2(u) du$. Define $\mathbf{M}$ to be a 2×2 matrix with elements $\mathbf{M}_{kl} = \kappa_{k+l-2}$ and $\mathbf{R}$ to be a 2×2 matrix with elements $\mathbf{R}_{kl} = \nu_{k+l-2}$. Let $\rho(\mathbf{x}, t) = 1/V \{ \mu(\mathbf{x}, t) \} g^2 \{ \mu(\mathbf{x}, t) \}$ and $\gamma(\mathbf{x}, t_0) = \text{var}(Y_1 / \mathbf{X}_1 = \mathbf{x}, T_1 = t_0) / [V \{ \mu(\mathbf{x}, t_0) \} g^4 \{ \mu(\mathbf{x}, t_0) \}]^2$. Make the definitions

$$\begin{aligned} \mathbf{V}_1(t_0) &= f_t(t_0) \mathbf{M} \otimes E\{\rho(\mathbf{X}_1, T_1) \mathbf{X}_1 \mathbf{X}_1^T | T_1 = t_0\}, \\ \mathbf{V}_2(t_0) &= f_t(t_0) \mathbf{R} \otimes E\{\gamma(\mathbf{X}_1, T_1) \mathbf{X}_1 \mathbf{X}_1^T | T_1 = t_0\}, \\ \mathbf{B}_1(t_0) &= \mathbf{M}^{-1}(\kappa_2, \kappa_3)^T \otimes \boldsymbol{\eta}''(t_0)/2, \end{aligned}$$

where $f_t(\cdot)$ is the marginal density function of T . Then for a fixed guide, we have the following result.

Theorem 2.1: Fix a point t_0 and assume the guide is fixed. Under the conditions stated in Section A.1 of the Appendix, as $h \rightarrow 0$, $nh \rightarrow \infty$, and $nh^5 \rightarrow \text{constant}$, we have

$$(nh)^{1/2} \{\hat{\boldsymbol{\theta}}(t_0) - \boldsymbol{\theta}(t_0) - h^2 \mathbf{B}_1(t_0)\} \rightarrow^d N(\mathbf{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma}$ is the leading $(q+1) \times (q+1)$ submatrix of $\mathbf{V}_1^{-1} \mathbf{V}_2 \mathbf{V}_1^{-1}$.

There are two noteworthy points to make about Theorem 1. The first is that if there is no (or constant) guide and the model belongs to a one-parameter exponential family with the canonical link and correctly specified variance function, then $\boldsymbol{\eta}''(t_0) = \boldsymbol{\theta}''(t_0)$, $\rho(\mathbf{x}, t) = \gamma(\mathbf{x}, t) = V\{\mu(\mathbf{x}, t)\}$, and our result reduces to Theorem 1 of Cai et al. (2000). The second is that only the bias term $\mathbf{B}_1(t_0)$ is affected when a guide is used, and not the asymptotic variance. To be specific, consider the integrated squared bias for each function $\theta_j(\cdot)$. It is evident that the integrated squared bias of the guided estimate is smaller than that of the unguided one when $\int \{\theta_j''(t) - \theta_{jg}''(t)\}^2 dt < \int \{\theta_j''(t)\}^2 dt$. This is analogous to Remark 3 of Fan et al. (2009). Therefore when a fixed bandwidth is used, the squared bias (and hence the MSE) is reduced if an appropriate guide is chosen. Furthermore, finding the optimal bandwidth using the procedure from Section 3 allows for a larger bandwidth to be selected compared to the unguided estimate which will reduce the variance as well as the bias. We demonstrate this in our simulation study in Section 4.

Theorem 2.1 assumes a fixed guide but in practice the guide needs to be estimated. Theorem 2.2 states that the expressions above are still valid for estimated guides as well.

Theorem 2.2: Let $f_{jnt}(\mathbf{x}, t, y) = f_t(t) \exp(Q[g^{-1}\{\mathbf{x}^T \boldsymbol{\theta}(t)\}, y])$ be the true joint density and $f_{prp}(\mathbf{x}, t, y, \boldsymbol{\alpha}) = f_t(t) \exp(Q[g^{-1}\{\mathbf{x}^T \boldsymbol{\theta}_g(t, \boldsymbol{\alpha})\}, y])$ be the proposed joint density. Define $\boldsymbol{\alpha}^*$ to be the minimizer of the Kullback-Leibler distance between f_{prp} and f_{jnt} . Then under the White (1982)-type conditions in the Appendix, the same result as in Theorem 2.1 holds when an estimated guide $\boldsymbol{\theta}_g(\cdot)$ is used in place of a fixed guide with the modification that $\boldsymbol{\alpha}$ is now replaced by $\boldsymbol{\alpha}^*$.

The proofs of Theorems 2.1 and 2.2 are given in the Appendix.

Remark 1: The asymptotic results presented in this section are an extension of the results from Fan et al. (2009). Consider the special case where the covariate contains only the intercept $\mathbf{X}_i = 1$ and parameter function $\boldsymbol{\theta}(\cdot) \equiv \theta_0(\cdot)$. Then we have $\rho(\mathbf{x}, t) = \rho(t) = 1/V$

$\{\mu(t)\}g'^2\{\mu(t)\}$, $\gamma(\mathbf{x}, t) = \gamma(t) = \text{var}(Y_1/T_1 = t)/[V\{\mu(t)\}g'\{\mu(t)\}]^2$, $\mathbf{V}_1(t_0) = f_t(t_0)\rho(t_0)\mathbf{M}$, $\mathbf{V}_2(t_0) = f_t(t_0)\gamma(t_0)\mathbf{R}$, and $\mathbf{B}_1(t_0) = \mathbf{M}^{-1}(\kappa_2, \kappa_3)^T \eta''(t_0)/2$. With these simplified definitions, the result in Theorem 2.1 implies that

$$(nh)^{1/2}\{\hat{\theta}(t_0) - \theta(t_0) - h^2 \kappa_0 \kappa_2 \eta''(t_0)/2\} \rightarrow^d N(0, \sigma(t_0)),$$

where $\sigma(t_0) = v_0 \text{var}(Y_1/T_1 = t_0) g'^2\{\mu(t)\}/f_t(t_0)$. This result is as in Theorem 1 in Fan et al. (2009) with the assumption that $\kappa_0 = \int K(z) dz = 1$.

3. Optimal Bandwidth Selection

Note that for simplicity of presentation, we will consider our bandwidth selection method using the additive correction; the multiplicative correction follows easily by using the multiplicative definition of $\eta_j(\cdot)$, replacing $\mathbf{G}_i$ with $\mathbf{G}_i^*$, and omitting the offset term $\mathbf{X}_i^T \mathbf{h}(T_i)$.

Once $\hat{\beta}$ is obtained, the bias arises from the approximation term of the Taylor series expansions. Hence, using more terms in the series should theoretically produce less bias. Let

$r_j(T_i) = \eta_j(T_i) - \sum_{k=0}^P \eta_j^{(k)}(t_0)(T_i - t_0)^k/k!$ be the approximation error. If a higher order Taylor approximation is substituted for $\eta_j(T_i)$ in r_j , then

$r_j(T_i) \approx \sum_{k=1}^a \eta_j^{(p+k)}(t_0)(T_i - t_0)^{p+k}/(p+k)! \stackrel{\text{def}}{=} r_{ji}$. We then maximize the local quasi-likelihood including the approximation errors

$$Q^*(\beta) = \sum_{i=1}^n Q[g^{-1}\{\mathbf{G}_i^T \beta + \mathbf{X}_i^T \mathbf{h}(T_i) + \mathbf{X}_i^T \mathbf{r}_i\}, Y_i] \times K_h(T_i - t_0)$$

with respect to β where $\mathbf{r}_i = (r_{0i}, \dots, r_{qi})^T$. Define the maximizer as $\hat{\beta}^*$. The local quasi-likelihood $Q^*(\beta)$ is differentiated with respect to β to get the gradient vector

$$Q^{*'}(\beta) = \sum_{i=1}^n \frac{Y_i - g^{-1}\{\mathbf{G}_i^T \beta + \mathbf{X}_i^T \mathbf{h}(T_i) + \mathbf{X}_i^T \mathbf{r}_i\}}{V[g^{-1}\{\mathbf{G}_i^T \beta + \mathbf{X}_i^T \mathbf{h}(T_i) + \mathbf{X}_i^T \mathbf{r}_i\}]} (g^{-1})'(\mathbf{G}_i^T \beta + \mathbf{X}_i^T \mathbf{h}(T_i) + \mathbf{X}_i^T \mathbf{r}_i) \mathbf{G}_i K_h(T_i - t_0)$$

and the second derivative is taken to get the Hessian matrix $Q^{*''}(\beta)$. A Taylor series expansion is then applied to $Q^{*'}$ around $\hat{\beta}$ to get

$$Q^{*'}(\hat{\beta}^*) \approx Q^{*'}(\hat{\beta}) + Q^{*''}(\hat{\beta})(\hat{\beta}^* - \hat{\beta}) = 0$$

and thus, an approximation of the estimation bias is $\hat{\beta} - \hat{\beta}^* \approx \{Q^{*''}(\hat{\beta})\}^{-1} Q^{*'}(\hat{\beta})$.

To get an approximation of the variance, a Taylor series expansion of $Q'(\hat{\beta})$ is done about the true β , denoted by β^0 . Note that

$$0 = Q'(\hat{\beta}^*) \approx Q'(\beta^0) + Q''(\beta^0)(\hat{\beta} - \beta^0)$$

which implies $\hat{\beta} - \beta^0 \approx -\{Q''(\beta^0)\}^{-1}Q'(\beta^0)$, and the estimate of the conditional variance is

$$\text{var}(\hat{\beta} - \beta^0 | \mathbb{X}, \mathbb{T}) \approx \{Q''(\beta^0)\}^{-1} \text{var}\{Q'(\beta^0) | \mathbb{X}, \mathbb{T}\} \{Q''(\beta^0)\}^{-1}$$

where $Q''(\beta^0)$ can be approximated by $Q''(\hat{\beta})$. To approximate the variance term, note that

$$\begin{aligned} \text{var}\{Q'(\beta^0) | \mathbb{X}, \mathbb{T}\} &= \sum_{i=1}^n \text{var} \left(\frac{\partial}{\partial \beta} Q[g^{-1}\{\mathbf{G}_i^T \beta + \mathbf{X}_i^T \mathbf{h}(T_i)\}, Y_i | \mathbf{X}_i, T_i] K_h^2(T_i - t_0) \right) \\ &= \sum_{i=1}^n \text{var} \left(\frac{Y_i - g^{-1}\{\mathbf{G}_i^T \beta + \mathbf{X}_i^T \mathbf{h}(T_i)\}}{V[g^{-1}\{\mathbf{G}_i^T \beta + \mathbf{X}_i^T \mathbf{h}(T_i)\}]} (g^{-1})' \{\mathbf{G}_i^T \beta + \mathbf{X}_i^T \mathbf{h}(T_i)\} | \mathbf{X}_i, T_i \right) \times \mathbf{G}_i \mathbf{G}_i^T K_h^2(T_i - t_0) \\ &\approx \frac{[(g^{-1})' \{\mathbf{X}_i^T \boldsymbol{\theta}(t_0)\}]^2}{V[g^{-1}\{\mathbf{X}_i^T \boldsymbol{\theta}(t_0)\}]} \mathbf{G}_i \mathbf{G}_i^T K_h^2(T_i - t_0). \end{aligned}$$

The last approximation follows from the fact that T_i only has significant weight in the neighborhood of t_0 .

To compute the MSE, we denote the bias of $\hat{\theta}$ as $\mathbf{B}(t_0; h) = \{B_0(t_0; h), \dots, B_q(t_0; h)\}^T$ corresponding to the first $q+1$ components of $[Q^{*''}(\hat{\beta})]^{-1}Q^{*'}(\hat{\beta})$. The variance-covariance matrix $\mathbf{V}(t_0; h)$ of $\hat{\theta}$ is the first $(q+1) \times (q+1)$ submatrix of $\text{var}(\hat{\beta} - \beta^0 | \mathbf{x}, t)$ and the variance of the estimated VCM $\mathbf{X}_i^T \hat{\theta}$ is $\mathbf{X}_i^T \mathbf{V}(t_0; h) \mathbf{X}_i$. The conditional MSE of $\mathbf{X}^T \hat{\theta}$ given $\mathbf{X} = \mathbf{x}$ is

$$\text{MSE}(t_0; h) = \mathbf{x}^T \{\mathbf{B}(t_0; h) \mathbf{B}(t_0; h)^T + \mathbf{V}(t_0; h)\} \mathbf{x}.$$

The sample MSE is derived as

$$\begin{aligned} \hat{\text{MSE}}(t_0; h) &= \frac{1}{h} \sum_{i=1}^n \mathbf{X}_i^T \{\mathbf{B}(t_0; h) \mathbf{B}(t_0; h)^T + \mathbf{V}(t_0; h)\} \mathbf{X}_i \\ &= \text{trace} \left[\{\mathbf{B}(t_0; h) \mathbf{B}(t_0; h)^T + \mathbf{V}(t_0; h)\} n^{-1} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T \right]. \end{aligned}$$

We propose to choose h such that

$$\hat{h} = \arg\min_h \int \hat{\text{MSE}}(t; h) dt.$$

To summarize, we use a grid of t_0 's and a grid of candidate bandwidths. We fit the local quasi-likelihood for each t_0 and h candidate and calculate the extended residual squares criterion (ERSC) defined as $\text{ERSC}(\mathbf{x}, t; h) = \hat{\sigma}^2 \{1 + (p+1)/N\}$, where $\hat{\sigma}^2$ is the weighted residual sum of squares after fitting a local p th-order polynomial and N is the number of local data points (see (5.6) of Fan, Farnen, and Gijbels (1998) for details). We then sum the ERSCs over t_0 and the bandwidth with the lowest sum becomes the pilot bandwidth. Using this bandwidth, we fit a local quasi-likelihood for each t_0 to obtain $\hat{\beta}$. For a new set of candidate bandwidths, we fit the local quasi-likelihood using higher order Taylor series approximation and obtain $\hat{\beta}^*$, which is theoretically more accurate. We compute the bias and variance using the gradient and Hessian of the quasi-likelihood, compute the MSE and the candidate bandwidth with the lowest MSE is our optimal bandwidth.

4. Simulation

We conducted a simulation study to evaluate the performance of our estimators. We generated each observation (X_i, T_i, Y_i) by first simulating the covariates X_i and T_i from a uniform distribution. Then, given X_i and T_i , the conditional mean μ_i of the response was generated as

$$\mu_i = g^{-1} \{ \theta_0(T_i) + X_{i1} \theta_1(T_i) \}$$

where g is the canonical link. We used a grid of equally spaced values t_k for $k = 1, \dots, K = 100$ to estimate the two functions. A cubic guide $\theta_{0g}(t, \alpha_0) = \alpha_{01} + \alpha_{02}t + \alpha_{03}t^2 + \alpha_{04}t^3$ was used for estimating θ_0 and a quadratic guide $\theta_{1g}(t, \alpha_1) = \alpha_{11} + \alpha_{12}t + \alpha_{13}t^2$ was used for estimating θ_1 . We used local linear polynomial estimators with the Epanechnikov kernel weight, and for bias calculations, we chose the degree of the Taylor expansion to be $a = 1$.

For $R = 1000$ simulations, we generated the data and estimated the parameters for the cubic and quadratic guides as described above. To get the final estimates $\hat{\theta}_0$ and $\hat{\theta}_1$, we maximized the local quasi-likelihood with the appropriate distribution and canonical link, and Epanechnikov kernel weight. The design matrix of the local likelihood was constructed using both the additive and multiplicative corrections ($\mathbf{G}_i$ and $\mathbf{G}_i^*$, respectively). To find the optimal smoothing bandwidth, we first simulated 15 data sets and applied our methods from Section 3. We took the median of these 15 values as the optimal bandwidth and used that value as fixed for the 1000 simulations. We also compared the two methods using the optimal bandwidth from the unguided method as the fixed bandwidth.

Once the two functions were estimated, we computed the marginal squared bias, marginal variance, and marginal MSE of each. Define

$$B_{jk} = R^{-1} \sum_{r=1}^R [\hat{\theta}_{j,r}(t_k) - \theta_j(t_k)], V_{jk} = R^{-1} \sum_{r=1}^R [\hat{\theta}_{j,r}(t_k) - R^{-1} \sum_{r'=1}^R \hat{\theta}_{j,r'}(t_k)]^2, \text{ and}$$

$\text{MSE}_{jk} = B_{jk}^2 + V_{jk}$ where $j = 0, 1$ and r indexes the simulation. The average marginal

squared bias of θ_j is $B_j^2 = K^{-1} \sum_{k=1}^K B_{1k}^2$, the average marginal variance is

$$V_j = K^{-1} \sum_{k=1}^K V_{jk} \text{ and the average marginal MSE is } \text{MSE}_j = K^{-1} \sum_{k=1}^K \text{MSE}_{jk}. \text{ However}$$

instead of averaging over all k , we used the 10% trimmed mean. Tables A1 – A3 show the results of our simulations. “Best h ” is each estimation method’s optimal smoothing bandwidth obtained via the bias-variance tradeoff in Section 3. All values for squared bias, variance, and MSE in the table are multiplied by 100. “Same h ” refers to the fixed bandwidth obtain from the optimal bandwidth from the unguided estimators.

Example 1: Poisson Response

For the Poisson response, $n = 100$ or 200 covariates X_i and T_i were generated with $X_i \sim \text{Unif}[-1, 1]$ and $T_i \sim \text{Unif}[-2, 2]$. The true functions were $\theta_0(t) = \sin(\pi/2) + 4$ and $\theta_1(t) = \sin(\pi/4 - \pi/2)/2 + 1$. The response Y_i was generated from a $\text{Poisson}(\mu_i)$. We used a grid of $K = 100$ equally spaced values in $[-2, 2]$ for t_0 to estimate the two functions. Table A1 gives the (trimmed) average marginal squared bias, variance, and MSE for the two functions estimated by the original method using no guide and by our method using guided estimation with additive and multiplicative corrections. When the same bandwidth is used the guided estimation procedure reduces bias but has no effect on variance. When the optimal bandwidth is used, the guided estimates have lower bias and lower variance. This is because the guides account for much of the trend in the true curve and the nonparametric correction is flatter and easier to estimate, resulting in lower bias. As the sample size increased, we saw a reduction in bias, variance and MSE.

Example 2: Normal Response

For the Gaussian response, $n = 100$ or 200 covariates X_i and T_i were generated with $X_i \sim \text{Unif}[-1, 1]$ and $T_i \sim \text{Unif}[-2, 2]$. The true functions were $\theta_0(t) = \sin(\pi/2) - 2$ and $\theta_1(t) = 2 \sin(\pi/4 - \pi/2) + 3$. The response Y_i was generated from a $\text{Normal}(\mu_i, 1)$. Table A2 gives the (trimmed) average marginal squared bias, variance, and MSE for the two functions estimated by the original method using no guide and by our method using guided estimation with additive and multiplicative corrections. In this example, the guided estimates still have lower bias and variance when the optimal bandwidth is used. When the same bandwidth is used, the variance of the additive and multiplicative correction is slightly higher than the unguided estimates. The gains in bias reduction are counteracted by the higher variance so the MSE is approximately the same for both methods.

Example 3: Bernoulli Response

For the Bernoulli response, we chose a larger sample size of $n = 500$ since the estimation of the success probability is more difficult than estimating the mean in the Gaussian and Poisson case. The covariates X_i and T_i were generated with $X_i \sim \text{Unif}[1, 2]$ and $T_i \sim \text{Unif}[-1, 1]$. The true functions were $\theta_0(t) = \sin(\pi)/2 + 1$ and $\theta_1(t) = 0.7 \sin\{(t+1)\pi/2\} - 1$. The response Y_i was generated from a $\text{Bernoulli}(\mu_i)$. The results from this example are presented in Table A3. Using a multiplicative correction with Bernoulli data is very unstable due to the possibility of dividing by zero, thus this correction was not used. We see that using the guides reduces bias and variance when the optimal bandwidth is used, and reduces bias but has little effect on variance when a fixed bandwidth is used.

5. HIV Data Analysis

In the AIDS Clinical Trials Group (ACTG) Protocol 315, 48 individuals infected with HIV-1 were given potent antiviral medicine to evaluate the efficacy of treatment on the reduction of viral load (plasma HIV-1 RNA). The viral load was measured repeatedly over three months and 31 baseline covariates were measured for each individual. Details of this study can be found in Lederman et al. (1998) and Liang, Wu, and Carroll (2003). Of these 31 covariates, Wu and Wu (2002) identified those that were significant predictors and in this illustration, we chose two of these covariates that appeared frequently in their models, baseline viral load and baseline CD4+ counts. Viral load is the number of copies per milliliter and is measured on a logarithmic scale with base 10. CD4+ counts are the number of lymphocytes that are CD4+. The response was the change from baseline viral load measured at day 7. If an individual did not have a viral load measurement on day 7, then the preceding or following day measurement was used. The data for our illustration are presented in Figure A1. We would like to determine if baseline viral load and baseline CD4+ counts have an effect on the change from baseline viral load measurement while adjusting for interaction between them.

Using a grid of size of 25 for t_0 , the Epanechnikov kernel, and the Gaussian log-likelihood for Q , we fit the model in (1) and estimated θ_0 and θ_1 using the original unguided nonparametric method, represented by the solid red line in Figure A2. The pre-asymptotic bandwidth selector gave bandwidth 0.67. Based on the shape of these unguided estimates, we chose a cubic guide for our guided estimation of θ_0 and θ_1 and used the additive correction. The results from our fit are represented by the green dot-dashed lines in Figure A2 and the guides used are the black dashed line. The bandwidth for our method was also 0.67. Our estimated $\hat{\theta}_0$ had more curvature than the naive counterpart and followed the parametric guide very closely, suggesting that there is little model misspecification when using a cubic guide for θ_0 . For θ_1 , the naive estimate and our guided estimate had somewhat similar shapes, with the most difference in the left endpoints.

The point-wise bootstrap 95% confidence intervals are given in Figure A3. The confidence intervals for the guided estimates were slightly wider in the boundaries but overall were similar to those of the naive estimate. In Figure A3(c) and (d), the entire confidence interval for the functions contain zero. Recall that θ_1 is the slope function and this term models the interaction between the two covariates. This indicates that there is no interaction between baseline viral load and baseline CD4+ counts and the response can be adequately modeled as a cubic function of baseline viral load alone.

We also used the VCM to separately model the viral load after day 14, day 21, and day 28 with the same two covariates in order to compare the estimated functions on different days to day 7. Again, if an individual did not have a viral load measurement on these exact days, then the preceding or following day was used, and if all three days were missing, the individual was dropped from the analysis. This yielded 35 individuals for the day 14 analysis, 39 individuals for the day 21 analysis, and 38 individuals for the day 28 analysis. The estimated functions corresponding to these responses are given in Figure A4. The pre-asymptotic bandwidth selector gave bandwidths 0.53, 1.22 and 0.77 for the naive estimates

of day 14, 21, and 28 respectively. The bandwidth for the guided estimates were 0.63, 1.26, and 0.77 for day 14, 21, and 28 respectively. The shape of the slope function θ_1 changes drastically for the different days, but the entire confidence interval for all three θ_1 functions (not presented) contains zero. Thus CD4+ has a very different interaction effect on baseline viral load for different days, but the overall effect is not significant. The shape of the intercept function θ_0 was similar for days 14 and 21, and had more of a cubic shape for day 28. Similar to day 7, the parametrically guided estimates of θ_0 followed their respective guides very closely.

6. Discussion

In this paper, we used parametric guides to enhance the performance of nonparametric estimators of the parameter functions in varying coefficient models. We generalized to quasi-likelihoods since the true likelihood is often unavailable. We presented two ways of using the guide and estimated the corrected functions using local polynomial fitting. We developed the asymptotic properties of the guided estimators and a method of selecting the optimal bandwidth parameter of the kernel function. We conducted a simulation study to compare our guided estimators to their standard nonparametric equivalents, and found that the guided estimators had lower bias when a fixed bandwidth was used, and lower bias and variance when the optimal bandwidth was used. In general, even if the shape of the parameter function is not captured by the guide, the guided estimator will still have better bias than the unguided counterpart and the two will have similar variability.

In this paper we present the additive and multiplicative correction which are special cases of a unified family of guided estimators proposed by Fan et al. (2009). This work could be extended to include this unified family and its asymptotic properties. Other future work includes extending our methods to functional data where the covariates of interest are smooth functions and the response can be functional or scalar. The functional covariates correspond to unknown functional parameters that need to be estimated, and this can be done using our guided estimation scheme. This work can also be extended to multivariate unknown functions (e.g. $\theta(\mathbf{T}_i)$) by using an empirical basis expansion of the function and reducing it to a VCM.

Acknowledgments

Maity's research was supported by NIH grant R00 ES017744 and a NCSU Faculty Research and Professional Development (FRPD) grant. Wu's research was partially supported by NSF grant DMS-1055210 and NIH grant R01-CA149569. We also thank two anonymous referees, an anonymous Associate Editor and the editor for their constructive and helpful comments, which has improved the presentation of the paper.

References

- Brumback B, Rice JA. 'Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves' (with discussion). *Journal of the American Statistical Association*. 1998; 93:961–994.
- Cai Z, Fan J, Li R. Efficient Estimation and Inferences for Varying-coefficient Models. *Journal of the American Statistical Association*. 2000; 95:888–902.
- Cao Y, Lin H, Wu TZ, Yu Y. Penalized Spline Estimation for Functional Coefficient Regression Models. *Computational Statistics and Data Analysis*. 2010; 54:891–905. [PubMed: 21516260]

- Chen R, Tsay RS. Functional-coefficient Autoregressive Models. *Journal of the American Statistical Association*. 1993; 88:298–308.
- Cleveland, W.; Grosse, E.; Shyu, W. Local Regression Models. In: Chambers, J.; Hastie, T., editors. *Statistical Models in S*. 2. Pacific Grove, CA: Wadsworth and Brooks/Cole; 1991. p. 309–376.
- Eilers PHC, Marx BD. Flexible Smoothing with B-splines and Penalties. *Statistical Science*. 1996; 11:89–102.
- Fan J. Local Linear Regression Smoothers and their Minimax Efficiencies. *The Annals of Statistics*. 1993; 21:196–216.
- Fan J, Farman M, Gijbels I. Local Maximum Likelihood Estimation and Inference. *Journal of the Royal Statistical Society Series B*. 1998; 60:591–608.
- Fan, J.; Gijbels, I. *Local Polynomial Modelling and its Applications*. London: Chapman and Hall; 1996.
- Fan J, Maity A, Wang Y, Wu Y. Parametrically Guided Generalised Additive Models with Application to Mergers and Acquisitions Data. *Journal of Non-parametric Statistics*. 2013; 25:109–128.
- Fan J, Wu Y, Feng Y. Local Quasi-likelihood with a Parametric Guide. *The Annals of Statistics*. 2009; 37:4153–4183.
- Fan J, Zhang W. Statistical Estimation in Varying Coefficient Models. *The 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]
- Glad IK. Parametrically Guided Non-parametric Regression. *Scandinavian Journal of Statistics*. 1998; 25:649–668.
- Hastie T, Tibshirani R. Varying-coefficient Models. *Journal of the Royal Statistical Society Series B*. 1993; 55:757–796.
- Hjort NL, Glad IK. Nonparametric Density Estimation with a Parametric Start. *The Annals of Statistics*. 1995; 23:882–904.
- Holdeman JT. A Method for the Approximation of Functions Defined by Formal Series Expansions in Orthogonal Polynomials. *Mathematics of Computation*. 1969; 23:275–287.
- Hoover DR, Rice JA, Wu CO, Yang LP. Nonparametric Smoothing Estimates of Time-varying Coefficient Models with Longitudinal Data. *Biometrika*. 1998; 85:809–822.
- Huang JZ, Shen H. Functional Coefficient Regression Models for Nonlinear Time Series: A Polynomial Spline Approach. *Scandinavian Journal of Statistics*. 2004; 31:515–534.
- Lederman MM, Connick E, Landay A, Kuritzkes DR, Spritzler J, StClair M, Kotzin BL, Fox L, Chiozzi MH, Leonard JM, et al. Immunologic Responses Associated with 12 Weeks of Combination Antiretroviral Therapy Consisting of Zidovudine, Lamivudine, and Ritonavir: Results of AIDS Clinical Trials Group Protocol 315. *The Journal of Infectious Diseases*. 2002; 179:70–79.
- Liang H, Wu H, Carroll RJ. The Relationship between Virologic and Immunologic Responses in AIDS Clinical Research using Mixed-effects Varying-coefficient Models with Measurement Error. *Biostatistics*. 2003; 4:297–312. [PubMed: 12925523]
- Ma S, Yang L, Romero R, Cui Y. Varying Coefficient Model for Gene-environment Interaction: A Non-linear Look. *Bioinformatics*. 2011; 27:2119–2126. [PubMed: 21690105]
- Martins-Filho C, Mishra S, Ullah A. A Class of Improved Parametrically Guided Nonparametric Regression Estimators. *Econometric Reviews*. 2008; 27:542–573.
- Nelder JA, Wedderburn RWM. Generalized Linear Models. *Journal of the Royal Statistical Society Series A*. 1972; 135:370–384.
- Wedderburn RWM. Quasi-likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika*. 1974; 61:439–447.
- White H. Maximum Likelihood Estimation of Misspecified Models. *Econometrica*. 1982; 50:1–25.
- Wood, SN. *Generalized Additive Models: An Introduction with R*. Boca Raton, FL: Chapman and Hall/CRC; 2006.
- Wu H, Wu L. Identification of Significant Host Factors for HIV Dynamics Modelled by Non-linear Mixed-effects Models. *Statistics in Medicine*. 2002; 21:753–771. [PubMed: 11870815]

Appendix A. Appendix

A.1. Assumptions, Definitions and Facts

Recall that $\mathbf{X}_i = (1, X_{i1}, \dots, X_{iq})^T$ and that $\mathbf{G}_i = \{1, (T_i - t_0)\}^T \otimes \mathbf{X}_i$ for $P = 1$. Also recall that $\rho(\mathbf{x}, t) = 1/[V\{\mu(\mathbf{x}, t)\}g'^2\{\mu(\mathbf{x}, t)\}]$. Define $Q_k(u, v) = {}^kQ\{g^{-1}(u), v\}/u^k$. Since $Q(\cdot)$ is a quasi-likelihood function, $Q_k(u, v)$ is linear in v for each fixed u . Thus for fixed $\mathbf{X} = \mathbf{x}$, we have

$$\begin{aligned} Q_1\{\mathbf{x}^T\boldsymbol{\theta}(t_0), \mu(\mathbf{x}, t_0)\} &= 0, \\ Q_2\{\mathbf{x}^T\boldsymbol{\theta}(t_0), \mu(\mathbf{x}, t_0)\} &= -\rho(\mathbf{x}, t_0). \end{aligned}$$

The following assumptions that we use to prove Theorem 2.1 and 2.2 are well used in the nonparametric regression and nonparametric varying coefficient modeling literature, see for examples Cai et al. (2000) and Fan et al. (2009).

- The kernel $K(\cdot)$ is a symmetric positive bounded function with $[-1, 1]$ as support, and follows the property that $\int uK(u) du = 0$.
- $E(|\mathbf{X}|^3|T = t_0)$ is continuous at t_0 .
- $E(Y^4|\mathbf{X} = \mathbf{x}, T = t_0)$ is bounded in a neighborhood of t_0 .
- The function $Q_2(u, v) < 0$ for each u and v in the range of the response variable.
- The functions $f_i(\cdot)$, $\mathbf{V}_1(\cdot)$, $V\{\mu(\mathbf{x}, \cdot)\}$, $V'\{\mu(\mathbf{x}, \cdot)\}$ and $g''' \{\mu(\mathbf{x}, \cdot)\}$ are continuous at t_0 . Also, $f_i(t_0) > 0$ and $\mathbf{V}_1(t_0) > 0$.
- The functions $\theta_j''(\cdot)$ for $j = 0, \dots, q$ are continuous in a neighborhood around t_0 .

For our results to be valid for the case when the guide is estimated, we use the following White (1982)-type condition, see e.g., Fan et al. (2009). Recall that we denote the guides as $\theta_{jg}(t) = \theta_{jg}(t, \boldsymbol{\alpha})$ and similarly $\hat{\theta}_{jg}(t) = \theta_{jg}(t, \hat{\boldsymbol{\alpha}})$, omitting the dependence on $\boldsymbol{\alpha}$ to ease notation.

- We assume that $E[\log\{f_{\text{int}}(\mathbf{x}, t, y)\}]$ exists. Also, there exists a function $m(\mathbf{x}, t, y)$ such that $|f_{\text{prp}}(\mathbf{x}, t, y) - m(\mathbf{x}, t, y)|$ is small.
- $E[\log\{f_{\text{int}}(\mathbf{x}, t, y)\} - \log\{f_{\text{prp}}(\mathbf{x}, t, y)\}]$ has a unique minimizer $\boldsymbol{\alpha}^*$.
- Each element $(\boldsymbol{\alpha} / \boldsymbol{\alpha}) \log\{f_{\text{prp}}(\mathbf{x}, t, y)\}$ is continuously differentiable in $\boldsymbol{\alpha}$.
- There are functions $m_2(\mathbf{x}, t, y)$ and $m_3(\mathbf{x}, t, y)$ such that for any $\boldsymbol{\alpha}$, the absolute value of each element of $[(\boldsymbol{\alpha} / \boldsymbol{\alpha}) \log\{f_{\text{prp}}(\mathbf{x}, t, y)\}][(\boldsymbol{\alpha} / \boldsymbol{\alpha}) \log\{f_{\text{prp}}(\mathbf{x}, t, y)\}]^T$ is bounded by $m_2(\mathbf{x}, t, y)$, and that the absolute value of each element of $[(\boldsymbol{\alpha} / \boldsymbol{\alpha}) \log\{f_{\text{prp}}(\mathbf{x}, t, y)\}]^T$ is bounded by $m_3(\mathbf{x}, t, y)$. Also $E\{m_2(\mathbf{x}, t, y)\}$ and $E\{m_3(\mathbf{x}, t, y)\}$ exist.
- The matrix $E[(\boldsymbol{\alpha} / \boldsymbol{\alpha}) \log\{f_{\text{prp}}(\mathbf{x}, t, y)\}][(\boldsymbol{\alpha} / \boldsymbol{\alpha}) \log\{f_{\text{prp}}(\mathbf{x}, t, y)\}]^T$ is nonsingular at $\boldsymbol{\alpha}^*$ and that $\boldsymbol{\alpha}^*$ is a regular point of the matrix $E[(\boldsymbol{\alpha} / \boldsymbol{\alpha}) \log\{f_{\text{prp}}(\mathbf{x}, t, y)\}]$.

A.2. Proof of Theorem 2.1

We first prove the result with a fixed guide and then address the case when the guide is estimated.

Fix a point t_0 . Recall that $\eta_j(t) = \theta_j(t) - \theta_{jg}(t) + \theta_{jg}(t_0)$. Define $b_n = (nh)^{-1/2}$ and

$\boldsymbol{\delta} = b_n^{-1} [\beta_{00} - \theta_0(t_0), \dots, \beta_{q0} - \theta_q(t_0), h\{\beta_{01} - \eta'_0(t_0)\}, \dots, h\{\beta_{q1} - \eta'_q(t_0)\}]^T$. Then it is straightforward to see that $\hat{\boldsymbol{\delta}}$ minimizes

$$\sum_{i=1}^n Q[g^{-1}\{b_n \mathbf{G}_i^T \boldsymbol{\delta} + s(T_i, t_0) + \mathbf{X}_i^T \boldsymbol{\theta}_g(T_i)\}, Y_i] K\{(T_i - t_0)/h\}$$

as a function of $\boldsymbol{\delta}$, where $s(T_i, t_0) = \mathbf{X}_i^T \{\boldsymbol{\eta}(t_0) + (T_i - t_0) \boldsymbol{\eta}'(t_0)\}$. By Taylor's expansion we obtain

$$\begin{aligned} \mathcal{L}(\boldsymbol{\delta}) &\equiv \sum_{i=1}^n Q[g^{-1}\{b_n \mathbf{G}_i^T \boldsymbol{\delta} + s(T_i, t_0) + \mathbf{X}_i^T \boldsymbol{\theta}_g(T_i)\}, Y_i] K\{(T_i - t_0)/h\} \\ &= \mathcal{L}_0 + \mathcal{L}_1^T \boldsymbol{\delta} + \boldsymbol{\delta}^T \mathcal{L}_2 \boldsymbol{\delta} / 2 + \mathcal{L}_3(\boldsymbol{\delta}), \end{aligned}$$

where

$$\begin{aligned} \mathcal{L}_0 &= \sum_{i=1}^n Q[g^{-1}\{s(T_i, t_0) + \mathbf{X}_i^T \boldsymbol{\theta}_g(T_i)\}, Y_i] K\{(T_i - t_0)/h\}; \\ \mathcal{L}_1 &= b_n \sum_{i=1}^n Q_1[\{s(T_i, t_0) + \mathbf{X}_i^T \boldsymbol{\theta}_g(T_i)\}, Y_i] K\{(T_i - t_0)/h\} \mathbf{G}_i; \\ \mathcal{L}_2 &= b_n^2 \sum_{i=1}^n Q_2[\{s(T_i, t_0) + \mathbf{X}_i^T \boldsymbol{\theta}_g(T_i)\}, Y_i] K\{(T_i - t_0)/h\} \mathbf{G}_i \mathbf{G}_i^T; \\ \mathcal{L}_3(\boldsymbol{\delta}) &= (b_n^3/6) \sum_{i=1}^n Q_3[\{s_i^* + \mathbf{X}_i^T \boldsymbol{\theta}_g(T_i)\}, Y_i] K\{(T_i - t_0)/h\} (\mathbf{G}_i^T \boldsymbol{\delta})^3 \end{aligned}$$

with s_i^* lying between $s(T_i, t_0)$ and $b_n \mathbf{G}_i^T \boldsymbol{\delta} + s(T_i, t_0)$.

We analyze each term separately. First we note that using similar argument as in Cai et al. (2000) and Fan et al. (2009), $E(|\mathcal{L}_3(\boldsymbol{\delta})|)$ is bounded by a term of the order

$$O(n b_n^3 E|Q_3(s_i^*, Y_1)(\mathbf{X}_1^T \boldsymbol{\delta})^3 K\{(T_i - t_0)/h\}|) = O(b_n). \quad (\text{A1})$$

The previous line makes use of the assumption that $K(\cdot)$ is bounded, $Q_3(u, v)$ is linear in v , and that $E(|\mathbf{X}|^3 | T = t_0)$ is continuous at t_0 .

Next, we take up $\mathcal{L}_2$. For each T_i such that $|T_i - t_0| < h$, a Taylor's series expansion gives us

$$\mathbf{X}_i^T \boldsymbol{\eta}(T_i) = s(T_i, t_0) + (T_i - t_0)^2 \mathbf{X}_i^T \boldsymbol{\eta}''(t_0)/2 + o_p(h^2). \quad (\text{A2})$$

Thus we have

$$Q_2[\{s(T_i, t_0) + \mathbf{X}_i^T \boldsymbol{\theta}_g(t_i)\}, \mu(\mathbf{X}_i, T_i)] = Q_2[\mathbf{X}_i^T \boldsymbol{\theta}(T_i), \mu(\mathbf{X}_i, T_i)] + o_p(1) = -\rho(\mathbf{X}_i, T_i) + o_p(1).$$

The expected value of the (j, k) th element of $\mathcal{L}_2$, with $1 \leq j, k \leq (q+1)$ is

$$\begin{aligned} E(\mathcal{L}_{2,jk}) &= h^{-1} \int Q_2[\{s(T_1, t_0) + \mathbf{X}_1^T \boldsymbol{\theta}_g(T_1)\}, \mu(\mathbf{X}_1, T_1)] X_{1j} X_{1k} \times K\{(T_1 - t_0)/h\} f_{\mathbf{X},t}(\mathbf{X}_1, T_1) dT_1 d\mathbf{X}_1 \\ &\quad \int Q_2[\{s(wh + t_0, t_0) + \mathbf{X}_1^T \boldsymbol{\theta}_g(wh + t_0)\}, \mu(\mathbf{X}_1, wh + t_0)] X_{1j} X_{1k} \times K(w) f_{\mathbf{X},t}(\mathbf{X}_1, wh + t_0) dw d\mathbf{X}_1 \\ &= -\int \rho(\mathbf{X}_1, wh + t_0) X_{1j} X_{1k} K(w) f_{\mathbf{X},t}(\mathbf{X}_1, wh + t_0) dw d\mathbf{X}_1 + o_p(1) \\ &= -f_t(t_0) E[\rho(\mathbf{X}_1, T_1) X_{1j} X_{1k} | T_1 = t_0] + o_p(1). \end{aligned}$$

Similarly, it is fairly straightforward to derive that

$$E(\mathcal{L}_2) = -f_t(t_0) \mathbf{M} \otimes E[\rho(\mathbf{X}_1, T_1) \mathbf{X}_1 \mathbf{X}_1^T | T_1 = t_0] + o(1),$$

where $\mathbf{M}$ is a 2×2 matrix with elements $\mathbf{M}_{kl} = \kappa_{k+l-2}$. Using similar calculations, we can show that $\text{var}(\mathcal{L}_{2,jk}) = O\{(nh)^{-1}\}$. Thus we have that

$$\mathcal{L}_2 = -f_t(t_0) \mathbf{M} \otimes E[\rho(\mathbf{X}_1, T_1) \mathbf{X}_1 \mathbf{X}_1^T | T_1 = t_0] + o_p(1) = -\mathbf{V}_1(t_0) + o_p(1). \quad (\text{A3})$$

Hence we can write

$$\mathcal{L}(\boldsymbol{\delta}) - \mathcal{L}_0 = \mathcal{L}_1^T \boldsymbol{\delta} - \boldsymbol{\delta}^T \mathbf{V}_1(t_0) \boldsymbol{\delta} + o_p(1).$$

Note that $\hat{\boldsymbol{\delta}}$ actually minimizes $\mathcal{L} - \mathcal{L}_0$ since $\mathcal{L}_0$ is free of $\boldsymbol{\delta}$. Now we use the quadratic approximation lemma (Fan and Gijbels 1996) and derive that

$$\hat{\boldsymbol{\delta}} = \mathbf{V}_1^{-1} \mathcal{L}_1 + o_p(1),$$

provided that $\mathcal{L}_1$ is a sequence of stochastically bounded vectors.

To prove the asymptotic normality of $\hat{\boldsymbol{\delta}}$ it is sufficient to show the same for $\mathcal{L}_1$. Note that using (A2) we have

$$Q_1[\{s(T_i, t_0) + \mathbf{X}_i^T \boldsymbol{\theta}_g(T_i)\}, \mu(\mathbf{X}_i, T_i)] = \rho(\mathbf{X}_i, T_i) (T_i - t_0)^2 \mathbf{X}_i^T \boldsymbol{\eta}''(t_0)/2 + o_p(h^2).$$

Denote the joint density of $\mathbf{X}$ and T by $f_{\mathbf{x},t}(\cdot, \cdot)$. Using the assumption that $nh^5 \rightarrow \text{constant}$ we derive for $1 \leq j \leq q+1$,

$$\begin{aligned} E(\mathcal{L}_{1,j}) &= nb_n \int Q_1[\{s(T_1, t_0) + \mathbf{X}_1^T \boldsymbol{\theta}_g(T_1)\}, \mu(\mathbf{X}_1, T_1)] K\{(T_1 - t_0)/h\} \times X_{1j} f_{\mathbf{x},t}(\mathbf{X}_1, T_1) d\mathbf{X}_1 dT_1 \\ &= (nh)^{1/2} \int Q_1[\{s(wh + t_0, t_0) + \mathbf{X}_1^T \boldsymbol{\theta}_g(wh + t_0)\}, \mu(\mathbf{X}_1, wh + t_0)] \times K(w) X_{1j} f_{\mathbf{x},t}(\mathbf{X}_1, wh + t_0) d\mathbf{X}_1 dw \\ &= \{(nh^5)^{1/2}/2\} \int \rho(\mathbf{X}_1, wh + t_0) w^2 \mathbf{X}_1^T \boldsymbol{\eta}''(t_0) \times K(w) X_{1j} f_{\mathbf{x},t}(\mathbf{X}_1, wh + t_0) d\mathbf{X}_1 dT_1 + o_p\{(nh^5)^{1/2}\} \\ &= \{(nh^5)^{1/2}/2\} \kappa_2 f_t(t_0) E[\mathbf{X}_1^T \boldsymbol{\eta}''(t_0) \rho(\mathbf{X}_1, T_1) X_{1j} | T_1 = t_0] + o_p\{(nh^5)^{1/2}\}. \end{aligned}$$

Similarly we derive that

$$E(\mathcal{L}_{1,j}) = \{(nh^5)^{1/2}/2\} \kappa_3 f_t(t_0) E[\mathbf{X}_1^T \boldsymbol{\eta}''(t_0) \rho(\mathbf{X}_1, T_1) X_{1j} | T_1 = t_0] + o_p\{(nh^5)^{1/2}\} \text{ for } q+2 \leq j \leq 2(q+1). \text{ Hence we have that}$$

$$\begin{aligned} E(\mathcal{L}_1) &= (nh^5)^{1/2} \{f_t(t_0)/2\} \begin{pmatrix} \kappa_2 \\ \kappa_3 \end{pmatrix} \otimes E[\rho(\mathbf{X}_1, T_1) \mathbf{X}_1 \mathbf{X}_1^T | T_1 = t_0] \boldsymbol{\eta}''(t_0) + o_p\{(nh^5)^{1/2}\} \\ &= (nh^5)^{1/2} \mathbf{B}_1(t_0) + o_p(1). \end{aligned}$$

Now we take up $\text{var}(\mathcal{L}_1)$. Define $\mathbf{Z}_i$ to be the term inside the sum in the definition of $\mathcal{L}_1$.

Since for any j, k , $\mathbf{Z}_j$ is independent of $\mathbf{Z}_k$ when $j \neq k$, it is straightforward to derive

$$\text{cov}(\mathcal{L}_{1,j}, \mathcal{L}_{1,k}) = nb_n^2 [E(Z_{1j} Z_{1k}) - E(Z_{1j}) E(Z_{1k})].$$

Using similar argument to those in the computation of $E(\mathcal{L}_1)$, we derive that $E(Z_{1j}) = O_p(h^3)$

for any j and thus $nb_n^2 E(Z_{1j}) E(Z_{1k}) = o_p(1)$. Next, for $1 \leq j, k \leq (q+1)$,

$$\begin{aligned} nb_n^2 E(Z_{1j} Z_{1k}) &= nb_n^2 E[Q_1^2[\{s(T_1, t_0) + \mathbf{X}_1^T \boldsymbol{\theta}_g(T_1)\}, Y_1] K^2\{(T_1 - t_0)/h\} X_{1j} X_{1k}] \\ &= \nu_0 f_t(t_0) E\{\gamma(\mathbf{X}_1, T_1) X_{1j} X_{1k} | T_1 = t_0\} + o_p(1), \end{aligned}$$

where $\gamma(\mathbf{x}, t_0) = \text{var}(Y_1 | \mathbf{X} = \mathbf{x}, T_1 = t_0) / [V\{\mu(\mathbf{x}, t_0)\} g'\{\mu(\mathbf{x}, t_0)\}]^2$. Similarly we obtain

$$\text{var}(\mathcal{L}_1) = f_t(t_0) \mathbf{R} \otimes E\{\gamma(\mathbf{X}_1, T_1) \mathbf{X}_1 \mathbf{X}_1^T | T_1 = t_0\} + o_p(1) = \mathbf{V}_2(t_0) + o_p(1), \quad (\text{A4})$$

where $\mathbf{R}$ is a 2×2 matrix with elements $\mathbf{R}_{jk} = \nu_{j+k-2}$.

The final step of the argument is to use the Cramer-Wold device, that is to show that for any vector $\mathbf{d}$

$$\{\mathbf{d}^T \text{var}(\mathcal{L}_1) \mathbf{d}\}^{-1/2} \{\mathbf{d}^T \mathcal{L}_1 - \mathbf{d}^T E(\mathcal{L}_1)\} \rightarrow^d N(0, 1).$$

We only need to check Lyapounov's condition. To this end, recall that $\mathcal{L}_1 = b_n \sum_{i=1}^n \mathbf{Z}_i$. It is sufficient to show that $nb_n^3 E(|\mathbf{d}^T \mathbf{Z}_1|^3) \rightarrow 0$. This result follows from a similar argument as that in (A1). Hence we obtain $\text{var}(\mathcal{L}_1)^{-1/2} \{ \mathcal{L}_1 - E(\mathcal{L}_1) \} \rightarrow N(0, 1)$.

Combining (A3) – (A4) we obtain

$$\hat{\delta} - (nh^5)^{1/2} \mathbf{V}^{-1}(t_0) \mathbf{B}_1(t_0) \rightarrow^p N(0, \mathbf{V}_1^{-1} \mathbf{V}_2 \mathbf{V}_1^{-1})$$

Theorem 2.1 now follows from the definition of $\hat{\delta}$.

A.3. Proof of Theorem 2.2

Recall the definition of $Q(\boldsymbol{\beta}; h, t, \boldsymbol{\alpha})$ from Section 2.2 for various guides. Denote $\hat{\boldsymbol{\beta}}(t, \hat{\boldsymbol{\alpha}})$ to be the maximizer of $Q(\boldsymbol{\beta}; h, t, \boldsymbol{\alpha})$. When an estimated guide $\hat{\boldsymbol{\alpha}}$ is used, it suffices to show that $\|\hat{\boldsymbol{\beta}}(t, \hat{\boldsymbol{\alpha}}) - \hat{\boldsymbol{\beta}}(t, \boldsymbol{\alpha}^*)\|$ converges to zero in probability with a rate faster than $(nh)^{1/2}$, the convergence rate shown in Theorem 2.1 with a fixed guide.

We start by observing two facts. First, $n^{-1/2} \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*\| = O_p(1)$, which can be shown using the second set of assumptions at the end of Section A.1. Second, using the assumption that $Q_2(u, v) < 0$ for each u and v in the range of the response variable, it is evident that both $Q(\boldsymbol{\beta}; h, t, \hat{\boldsymbol{\alpha}})$ and $Q(\boldsymbol{\beta}; h, t, \boldsymbol{\alpha}^*)$ are strictly concave in $\boldsymbol{\beta}$. Thus we have

$$n^{-1} Q(\boldsymbol{\beta}; h, t, \hat{\boldsymbol{\alpha}}) = n^{-1} Q(\boldsymbol{\beta}; h, t, \boldsymbol{\alpha}^*) + O_p(n^{-1/2}); \quad (\text{A5})$$

$$n^{-1} \frac{\partial}{\partial \boldsymbol{\beta}} Q(\boldsymbol{\beta}; h, t, \hat{\boldsymbol{\alpha}}) = n^{-1} \frac{\partial}{\partial \boldsymbol{\beta}} Q(\boldsymbol{\beta}; h, t, \boldsymbol{\alpha}^*) + O_p(n^{-1/2}); \quad (\text{A6})$$

$$n^{-1} \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} Q(\boldsymbol{\beta}; h, t, \hat{\boldsymbol{\alpha}}) = n^{-1} \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} Q(\boldsymbol{\beta}; h, t, \boldsymbol{\alpha}^*) + O_p(n^{-1/2}), \quad (\text{A7})$$

for any $\boldsymbol{\beta}$, where the last two equations hold entry-wise. Thus it follows from (A5) that $\|\hat{\boldsymbol{\beta}}(t, \hat{\boldsymbol{\alpha}}) - \hat{\boldsymbol{\beta}}(t, \boldsymbol{\alpha}^*)\| = o_p(1)$.

Using (A6), and a Taylor's expansion, we have

$$\begin{aligned} 0 &= n^{-1} \frac{\partial}{\partial \boldsymbol{\beta}} Q(\boldsymbol{\beta}; h, t, \hat{\boldsymbol{\alpha}}) \Big|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}(t, \hat{\boldsymbol{\alpha}})} \\ &= n^{-1} \frac{\partial}{\partial \boldsymbol{\beta}} Q(\boldsymbol{\beta}; h, t, \boldsymbol{\alpha}^*) \Big|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}(t, \hat{\boldsymbol{\alpha}})} + O_p(n^{-1/2}) \\ &= n^{-1} \frac{\partial}{\partial \boldsymbol{\beta}} Q(\boldsymbol{\beta}; h, t, \boldsymbol{\alpha}^*) \Big|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}(t, \boldsymbol{\alpha}^*)} + n^{-1} \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} Q(\boldsymbol{\beta}; h, t, \boldsymbol{\alpha}^*) \Big|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}(t, \boldsymbol{\alpha}^*)} [\hat{\boldsymbol{\beta}}(t, \hat{\boldsymbol{\alpha}}) - \hat{\boldsymbol{\beta}}(t, \boldsymbol{\alpha}^*)] + O_p(n^{-1/2}). \end{aligned}$$

Recall that by definition of $\hat{\beta}(t, \alpha^*)$ we have $\frac{\partial}{\partial \beta} Q(\beta; h, t, \alpha^*)|_{\beta=\hat{\beta}(t, \alpha^*)}=0$. Also, using similar argument used to derive (A3), the Frobenius norm of the second derivative term in the expansion above can be shown to be $O_p(1)$. Thus we have that

$$\|\hat{\beta}(t, \hat{\alpha}) - \hat{\beta}(t, \alpha^*)\| = O_p(n^{-1/2}).$$

The rate $n^{1/2}$ is much faster than the rate $(nh)^{1/2}$ shown in Theorem 2.1, and hence the result follows.

Figure A1.

Scatterplot of the HIV Data variables of interest presented in Section 5. The response is change from baseline viral load at day 7.

Figure A2.

Nonparametric estimate (solid red) and parametrically guided estimate (green dot-dashed) of θ_0 (a) and θ_1 (b) along with the cubic guide (black dashed).

Figure A3.

Nonparametric estimate (solid red) of θ_0 (a) and θ_1 (c) and parametrically guided estimate (solid green) of θ_1 (b) and θ_2 (d) along with a point-wise 95% bootstrap confidence intervals (black dashed).

Figure A4.

Nonparametric estimate (solid red) and parametrically guided estimate (green dot-dashed) of θ_0 (a)–(c) and θ_1 (d)–(f) for day 14, day 21, and day 28 responses. A cubic guide was used in (a), (c), and (f), a quadratic guide was used in (b) and (e), and a quartic guide was in (d).

Table A1

Results of trimmed average bias, variance, and MSE for Example 1. “Same h ” refers to fixed bandwidth and best “ h ” refers to optimal bandwidth. All values are multiplied by 100.

		Bias ²			Variance			MSE		
		Naive	Additive	Multiplicative	Naive	Additive	Multiplicative	Naive	Additive	Multiplicative
$n = 100$	Same h	$\theta_0(t)$	0.129	0.022	0.022	0.179	0.173	0.174	0.308	0.195
		$\theta_1(t)$	0.003	0.001	0.001	0.552	0.522	0.522	0.556	0.522
	Best h	$\theta_0(t)$	0.129	0.098	0.096	0.179	0.104	0.106	0.308	0.203
		$\theta_1(t)$	0.003	0.001	0.001	0.552	0.294	0.296	0.556	0.295
	Same h	$\theta_0(t)$	0.080	0.014	0.015	0.090	0.088	0.089	0.171	0.103
		$\theta_1(t)$	0.003	0.001	0.001	0.269	0.261	0.262	0.272	0.262
$n = 200$	Best h	$\theta_0(t)$	0.080	0.026	0.026	0.090	0.074	0.075	0.171	0.100
		$\theta_1(t)$	0.003	0.001	0.001	0.269	0.217	0.217	0.272	0.217

Table A2

Results of trimmed average bias, variance, and MSE for Example 2. All values are multiplied by 100.

		Bias ²			Variance			MSE			
		Naive	Additive	Multiplicative	Naive	Additive	Multiplicative	Additive	Multiplicative		
<i>n</i> = 100	Same <i>h</i>	$\theta_0(t)$	1.05	0.11	0.10	3.82	4.15	4.29	4.86	4.26	4.39
		$\theta_1(t)$	0.20	0.02	0.02	11.68	11.81	12.21	11.89	11.83	12.23
	Best <i>h</i>	$\theta_0(t)$	1.05	0.22	0.21	3.82	3.41	3.51	4.86	3.63	3.73
		$\theta_1(t)$	0.20	0.04	0.03	11.68	8.73	9.92	11.89	8.77	8.96
	Same <i>h</i>	$\theta_0(t)$	0.46	0.07	0.07	2.20	2.31	2.38	2.66	2.38	2.45
		$\theta_1(t)$	0.13	0.01	0.01	6.65	6.70	6.85	6.78	6.71	6.85
<i>n</i> = 200	Best <i>h</i>	$\theta_0(t)$	0.46	0.11	0.11	2.20	2.00	2.05	2.66	2.11	2.16
		$\theta_1(t)$	0.13	0.01	0.01	6.65	5.55	5.63	6.78	5.56	5.64

Table A3

Results of trimmed average bias, variance, and MSE for Example 3. All values are multiplied by 100.

		Bias ²		Variance		MSE	
		Naive	Additive	Naive	Additive	Naive	Additive
n=500	Same h	$\theta_0(r)$	0.84	0.17	62.43	67.23	63.27
		$\theta_1(r)$	0.21	0.02	27.10	29.08	27.31
	Best h	$\theta_0(r)$	0.84	0.18	62.34	57.60	63.27
		$\theta_1(r)$	0.21	0.02	27.10	24.95	27.31