Published in final edited form as:

Stat Probab Lett. 2012 October ; 82(10): 1807–1814. doi:10.1016/j.spl.2012.06.002.

Partially Linear Varying Coefficient Models Stratified by a Functional Covariate

Arnab Maity and

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

Jianhua Z. Huang

Department of Statistics, Texas A&M University, College Station TX 77843-3143, U.S.A

Arnab Maity: amaity@ncsu.edu; Jianhua Z. Huang: jianhua@stat.tamu.edu

Abstract

We consider the problem of estimation in semiparametric varying coefficient models where the covariate modifying the varying coefficients is functional and is modeled nonparametrically. We develop a kernel-based estimator of the nonparametric component and a profiling estimator of the parametric component of the model and derive their asymptotic properties. Specifically, we show the consistency of the nonparametric functional estimates and derive the asymptotic expansion of the estimates of the parametric component. We illustrate the performance of our methodology using a simulation study and a real data application.

Keywords

Functional regression; Kernel smoothing; Profile method; Semi-varying coefficient model

1 Introduction

Suppose we observe for $i = 1, \dots, n$ outcome Y_i and covariates X_i , S_i and $\mathcal{Z}_i$, where $X_i = (X_{i1}, \dots, X_{ip})^T \in \mathbb{R}^p$, $S_i = (S_{i1}, \dots, S_{iq})^T \in \mathbb{R}^q$ and $\mathcal{Z}_i$ is a functional covariate assumed to belong to a function space $\mathcal{H}$. We consider the following partially linear functional varying coefficient model:

$$Y_i = X_i^T \beta_0(\mathcal{Z}_i) + S_i^T \eta_0 + \varepsilon_i, \quad (1)$$

where $\beta_0(t) = [\beta_{01}(t), \dots, \beta_{0p}(t)]^T$ with $\beta_{0k}(\cdot)$, $k = 1, \dots, p$, assumed to be smooth but unknown functionals, $\eta_0 \in \mathbb{R}^q$ and the random errors ε_i are independently and identically distributed with $E[\varepsilon_i | X, \mathcal{Z}, S] = 0$. We also assume the functional covariate $\mathcal{Z}$ is observed on a finite grid $0 = t_1 < \dots < t_s = 1$, without loss of generality.

This model is the functional generalization of the standard partially linear varying coefficient model where $\mathcal{Z}$ is typically assumed to be a scalar or a vector (Ahmad et al. 2005). A simple example of such model is when $\mathcal{Z}$ is binary (taking values 0 or 1) indicating a non-exposed or exposed group to some treatment. In that case, $\beta_0(\mathcal{Z})$ has only two values

indicating the different effects of X for the two groups. In general, $\mathcal{Z}$ can be thought as a stratification variable that divides the population into several strata and the varying coefficient model assumes that X has different effect on Y in different strata. In other words, it is a means to quantify the interaction between X and $\mathcal{Z}$. Such varying coefficient models have obtained much attention in the past years due to their usefulness and a fair amount of literature exists when the nonparametrically modeled variable is a vector. For example, Fan and Huang (2005) considered semiparametric varying-coefficient partially linear models and derived asymptotic properties of profile likelihood estimators. For a much more detailed review, readers are directed to Fan and Zhang (2008) and the references therein. However, in this paper, we consider the case where $\mathcal{Z}$ is a functional stratifier and develop an estimation procedure for β_0 and η_0 .

Semiparametric models are not well studied when covariates modeled nonparametrically are functional-valued. There is some recent literature on partially linear functional models. Aneiros-Perez and Vieu (2006) considered a special case of model (1) with $X_j \equiv 1$. Aneiros-Perez and Vieu (2008) considered semi-functional partial linear time series modeling to use a continuous path in the past (used as the functional covariate) to predict future values of the process. Lian (2011) considered functional partial linear models in the sense that the functional covariate is modeled via a linear operator while the other covariate (assumed to be in a vectorial topological space with a semi-metric) is modeled nonparametrically. Ferraty and Romain (2010) also provided detailed discussion about functional partial linear models, their estimation procedure using kernel smoothing and asymptotic properties of the estimates. However, to the best of our knowledge, there is no results on partially linear functional varying-coefficient model where the varying coefficients are functionals of the observed functional covariates and are modeled nonparametrically.

In this note, we extend the work of Aneiros-Perez and Vieu (2006) and also discuss asymptotic expansion and properties of estimators of η_0 , which is new when there exist functional covariates that enter a regression model nonparametrically. Section 2 presents our estimation method and relevant asymptotic results. Some simulation results are presented in Section 3 and an application to real data is given in Section 4. The Appendix collects all technical conditions and proofs.

2 Estimation method and asymptotic results

2.1 Estimation of $\beta_0(\cdot)$

We first discuss about estimation of $\beta(\cdot)$ for any fixed value of $\eta_0 = \eta^*$. Recall that we assume the covariates $\mathcal{Z}$ to be in a abstract space $\mathcal{H}$. Ferraty and Vieu (2004, 2006) argued that the most interesting spaces for functional data modeling are the so called semi-metric spaces. Here, to keep the general formulation, we assume that $\mathcal{H}$ is a semi-metric space with an associated semi-metric $d(\cdot, \cdot)$.

We now use Nadaraya-Watson type local constant kernel based estimators for β . Let $K(\cdot)$ denote the kernel function and h be a bandwidth, and define $K_h(z) = K(z/h)$. In general, $K(\cdot)$ should satisfy some usual smoothness conditions, see Appendix for details. Define $\mathcal{Y}(\eta^*) = Y - S^T \eta^*$. Then, for any given $z \in \mathcal{H}$, we estimate $\beta_0(z)$ by minimizing

$$\sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} \{\mathcal{Y}_i(\eta^*) - X_i^T \beta(z)\}^2,$$

with respect to $\beta(\cdot)$ or equivalently, one solves for $\beta(z)$,

$$0 = \sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} X_i \{\mathcal{Y}_i(\eta^*) - X_i^T \beta(z)\}.$$

Of course, the solution has a closed form:

$$\widehat{\beta}(z, \eta^*) = \left[\sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} X_i X_i^T \right]^{-1} \sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} X_i \mathcal{Y}_i(\eta^*). \quad (2)$$

Define $\beta(z, \eta) = [E(XX^T / \mathcal{Z} = z)]^{-1} E[X(Y - S^T \eta) / \mathcal{Z} = z]$. Note that $\beta_0(z) = \beta(z, \eta_0)$. Suppose the functions $\mathcal{Z}_1, \dots, \mathcal{Z}_n \in \mathcal{C}$ where $\mathcal{C}$ is a compact subset of $\mathcal{H}$ such that condition (C.2.1) in the Appendix is satisfied. Then we have the following result.

Result 1—Under regularity conditions given in the Appendix, we have that uniformly in $z \in \mathcal{C}$

$$\widehat{\beta}(z, \eta_0) - \beta(z, \eta_0) = O(h^a) + O\left(\sqrt{\frac{\log n}{n\varphi(h)}}\right) a.s.,$$

where $\varphi(h)$ is defined in the Appendix.

The proof of Result 1 is given in the Appendix.

2.2 Estimation of η_0

We estimate η_0 using the profile method, as described below. Recall that for each value of η , we can construct $\widehat{\beta}(z, \eta)$ as in (2). Then we propose to estimate η_0 by minimizing

$$\sum_{i=1}^n \{Y_i - S_i^T \eta - X_i^T \widehat{\beta}(\mathcal{Z}_i, \eta)\}^2$$

and thereby solving for η

$$0 = n^{-1} \sum_{i=1}^n \{S_i + \widehat{\beta}_\eta^T(\mathcal{Z}_i) X_i\} \{Y_i - S_i^T \eta - X_i^T \widehat{\beta}(\mathcal{Z}_i, \eta)\},$$

where

$$\begin{aligned} \widehat{\beta}_\eta(z) &= [\widehat{\beta}_{\eta,1}(z), \dots, \widehat{\beta}_{\eta,q}(z)]^T; \\ \widehat{\beta}_{\eta,j}(z) &= - \left[\sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} X_i X_i^T \right]^{-1} \sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} X_i S_{ij}. \end{aligned}$$

To investigate the asymptotic properties of $\hat{\eta}$, we first define $\beta_{\eta,j}(z) = -[E(XX^T | Z = z)]^{-1} E[XS_j | Z = z]$, and $\tilde{S} = S - E[XS^T | \mathcal{Z}] [E(XX^T | \mathcal{Z})]^{-1} X = S + \beta_{\eta}^T(\mathcal{Z})X$. Then we have the following result.

Result 2—Under regularity conditions given in the Appendix, we have that uniformly in $z \in \mathcal{C}$

$$n^{1/2}(\hat{\eta} - \eta_0) = n^{-1/2} \sum_{i=1}^n \tilde{S}_i \varepsilon_i + o_p(1) \rightarrow \text{Normal}(0, \sigma^2 \Sigma^{-1}),$$

where $\sigma^2 = E(\varepsilon^2)$ and $\Sigma = E(\tilde{S}\tilde{S}^T)$.

Remark 1—Note that by definition, $\tilde{S}$ does not depend on the true value of η_0 , and hence Σ is also independent of η_0 . Also it is interesting to note that $\tilde{S}_j$, the j^{th} element of $\tilde{S}$ for each j , can be thought of the residual of a nonparametric varying coefficient model with S_j the j^{th} element of S , as response and X and Z as covariates. Specifically, if we posit the nonparametric varying coefficient model $S_{ij} = X_i^T \mathcal{G}_j(\mathcal{Z}_i) + \zeta_i$, $i = 1, \dots, n$, $E(\zeta_i | X_i, Z_i) = 0$, then by similar arguments as in Result 1, $\hat{\eta}_j(z)$ gives us consistent estimate of $-\beta_{\eta,j}(z)$, regardless of the correctness of the posited model. Thus $\tilde{S}$ can be estimated by the residuals of the element-wise regressions of S on X and Z using nonparametric varying coefficient models. Consequently, one can easily estimate Σ using plug-in method where one replaces expectation by the corresponding sample version in the expression of Σ . The error variance σ^2 can also be estimated by the mean square error of the estimated residuals from the full semiparametric model.

3 Simulation Study

In this section we present a simulation study to evaluate the performance of our method. We consider the model

$$Y_i = X_{i1}\beta_1(\mathcal{Z}_i) + X_{i2}\beta_2(\mathcal{Z}_i) + \eta_{00} + S_i\eta_{01} + \varepsilon_i,$$

where we set $X_{i1} = \text{Normal}(0, 1)$, $X_{i2} = \text{Normal}(0, 1)$, $S_i = \text{Normal}(0, 1)$, and $\varepsilon_i = \text{Normal}(0, 1)$. The true values of the parameters are $\eta_0 = (\eta_{00}, \eta_{01})^T = (5, 5)^T$. The functional covariates are taken to be of the form

$$\mathcal{Z}_i(t) = a_i(t - 0.5)^2 + b_i$$

with $a_i = \text{Normal}(0, 4)$, $b_i = \text{Normal}(0, 1)$ and $t \in [0, 1]$. We take for $\mathcal{Z}(t) \in \mathcal{H}$

$$\begin{aligned} \beta_1(z) &= \int_0^1 \{z^{(1)}(t)\}^2 dt; \\ \beta_2(z) &= \int_0^1 \{|z^{(1)}(t)|\} \log\{|z^{(1)}(t)|\} dt. \end{aligned}$$

The functions are observed in a grid of 51 equally spaced points on $[0, 1]$. We use B-splines to approximate the observed functions and use first derivative based semi-metric. To fit each model, we use k -nearest neighbors type bandwidths which are selected by cross-validation.

To evaluate the performance of the estimator, we set up a test set of functions $\mathcal{T} = \{a(t - 0.5)^2 + 1; -3 \leq a \leq 3\}$. We consider three different sample sizes, $n = 200, 400$ and 1000 , and generate 500 data sets for each case.

For each data set, we estimate $\beta_1(z)$ and $\beta_2(z)$ with $z \in \mathcal{T}$ and compare the estimates with the actual values. The results are shown in Figure 1. Plotted are the actual values of the operators with $z \in \mathcal{T}$ and the median of the estimates over all simulated data sets along with 95% point-wise confidence bands. It is evident from the graphs that the estimates are very close to the true functional values for $z \in \mathcal{T}$ with reasonable confidence bands for both the functionals.

Also, we estimate η_{00} and η_{01} for each of the settings. We compute, for $j = 0, 1$, the scaled empirical bias as average of $n^{1/2}(\hat{\eta}_{0j} - \eta_{0j})$, scaled empirical standard error as the standard deviation of $n^{1/2}(\hat{\eta}_{0j} - \eta_{0j})$, estimated standard error as the average of estimated standard deviation of $n^{1/2}(\hat{\eta}_{0j} - \eta_{0j})$ (using Remark 1) and empirical coverage at 95% nominal level. The results are given in Table 1. It is evident that as n increases, the bias remains close to zero, with reasonable coverage and standard errors. To observe normality of the estimates, we provide Q-Q plot of the estimates of η_{00} and η_{01} against normal quantiles when $n = 1000$ in Figure 2. It is evident that for large sample size, the distribution of estimates behaves like a Gaussian distribution.

4 Data Example

We demonstrate our method using the Tecator data set available at <http://lib.stat.cmu.edu/datasets/tecator>. The goal is to predict the fat content of a meat sample on the basis of its near infrared absorbance spectrum. Each sample contains finely chopped pure meat with different moisture, fat and protein contents, which are measured in percent and are determined by analytic chemistry. The functional covariate for each food sample consist of a 100 channel spectrum of absorbances recorded on a Tecator Infratec Food and Feed Analyzer working in the wavelength range 850–1050 nm by the near infrared transmission (NIT) principle. See Ferraty and Vieu (2006) for a detailed discussion of the data set.

In this analysis, we demonstrate our method using a subset of the full data. We have for $i = 1, \dots, 215$ samples where our response variable Y_i is the fat content in the sample. We denote the protein and moisture content by X_1 and X_2 , respectively, and denote the functional covariate by $\mathcal{Z}$. We consider various (varying coefficient) models. To evaluate the performance of the models, we divide the data set into a training data consisting of 172 samples and a test set with 43 samples as suggested in the data documentation given at the above mentioned link. We use the prediction mean squared error

$$PMSE = \sum_{i=1}^{43} (\hat{Y}_i - Y_i)^2 / 43$$

as our evaluation criterion. To fit each model, we use k -nearest neighbors type bandwidths which are selected by cross-validation over the training sample. We use derivative based semi-metrics $d_k(f, g) = [\int \{f^{(k)}(x) - g^{(k)}(x)\}^2 dx]^{1/2}$ for $k = 0, 1, 2$. The results are displayed in Table 2.

It is evident that the model with linear effect of X_1 and varying coefficient for X_2 performs best among the all considered models. It is interesting to note that Aneiros-Perez and Vieu (2006) found that the Semi-functional partial linear model with both X_1 and X_2 linear was best among their models. From our results, it appears that allowing X_2 to have a varying coefficient captured the data better. In fact, the model with both X_1 and X_2 having varying coefficient performs competitively as well.

We emphasize that the example provided in this section is not meant to be a full case study or analysis of the data, but a demonstrative example that shows the performance of our methods and illustrates the fact that the partially linear functional varying coefficient models are competitive for this kind of data sets.

Acknowledgments

Maity's work was partly supported by Award Number R00ES017744 from the National Institute of Environmental Health Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Environmental Health Sciences or the National Institutes of Health. Huang's work was partly supported by grants from NSF (DMS-09-07170 AND DMS-10-07618) and King Abdullah University of Science and Technology (KUS-CI-016-04). We are also grateful to two anonymous reviewers for their careful evaluation of the paper and constructive comments that led to a significantly improved version of the paper.

References

- Aneiros-Perez G, Vieu P. Semi-functional partial linear regression. *Statistics & Probability Letters*. 2006; 76:1102–1110.
- Aneiros-Perez G, Vieu P. Nonparametric time series prediction: A semi-functional partial linear modeling. *Journal of Multivariate Analysis*. 2008; 99:834–857.
- Ahmad I, Leelahanon S, Li Q. Efficient estimation of a semiparametric partially linear varying coefficient model. *Annals of Statistics*. 2005; 33:256–283.
- Fan J, Huang T. Profile Likelihood Inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*. 2005; 11:1031–1057.
- Fan J, Zhang W. Statistical methods with varying coefficient models. *Statistics and Its Interface*. 2008; 1:179–195. [PubMed: 18978950]
- Ferraty F, Laksaci A, Tadj A, Vieu P. Rate of uniform consistency for nonparametric estimates with functional variables. *Journal of Statistical Planning and Inference*. 2010; 140:335–352.
- Ferraty, F.; Romain, Y. *The Oxford Handbook of Functional Data Analysis*. Oxford University Press; 2011.
- Ferraty F, Vieu P. Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *Journal of Nonparametric Statistics*. 2004; 16:111–125.
- Ferraty, F.; Vieu, P. *Nonparametric functional data analysis*. Springer; New York: 2006.
- Ferraty F, Vieu P. Erratum of: 'Non-parametric models for functional data, with application in regression, time-series prediction and curve discrimination'. *Journal of Nonparametric Statistics*. 2008; 20:187–189.
- Lian H. Functional partial linear model. *Journal of Nonparametric Statistics*. 2011; 23:115–128.

A Appendix

A.1 Conditions and Assumptions

We use the following assumptions for the results stated in this paper. These assumptions are usual in nonparametric functional modeling, see for example Ferraty and Vieu (2006) and Aneiros-Perez and Vieu (2006).

- (C. Z1) The functions $z_1, \dots, z_n \in \mathcal{C}$ where $\mathcal{C}$ is a compact subset of $\mathcal{H}$ such that $\bigcup_{k=1}^n \mathcal{B}(z_k, \ell_n)$, with $\tau_n \ell_n^\gamma = c_0$ for some real positive constants γ, c_0 .
- (C.K) $K(\cdot)$ has support $[0, 1]$, is Lipschitz continuous and is strictly decreasing. Also, there exists $c_K > 0$ such that for all $t \in [0, 1]$, $-K'(t) > c_K$.
- (C. θ) There exist constants c and a so that $|f(z_1) - f(z_2)| \leq c d(z_1, z_2)^a$ for all $z_1, z_2 \in \mathcal{H}$ with $f \in \{\beta_1, \dots, \beta_p, \beta_{\eta,1}, \dots, \beta_{\eta,q}, g_{11}, \dots, g_{pp}, w_{11}, \dots, w_{pq}\}$, where $g_{jk}(z) = E(X_j | X_k = z)$ and $w_{jk}(z) = E(X_j | S_k = z)$.
- (C.Y) For some $\beta \geq 2$, $E|Y|^\beta < \infty$.
- (C. Z2) There exists a function $\varphi(h)$ with $\int_0^1 \varphi(ht) dt / \varphi(h) > \theta_2 > 0$ for some $\theta_2 > 0$ such that for all $z \in \mathcal{H}$, there are constants c_1 and c_2 such that
- $$c_1 \varphi(h) \leq P[\mathcal{Z} \in B(z, h)] \leq c_2 \varphi(h).$$
- (C.X) $E(XX^T | \mathcal{Z})$ is finite positive definite. Also for all $m \geq 2$, $E(|X_j X_k|^m | \mathcal{Z} = z) < \sigma_{1m}(z) < \infty$, where $\sigma_{1m}(\cdot)$ is continuous at z .
- (C.S) $E(SS^T | \mathcal{Z})$ is finite positive definite. Also for all $m \geq 2$, $E(|X_j S_k|^m | \mathcal{Z} = z) < \sigma_{2m}(z) < \infty$, where $\sigma_{2m}(\cdot)$ is continuous at z .
- (C.e) $\sigma^2 = E(\varepsilon^2) > 0$ and $E|\varepsilon|^r < \infty$ for some $r \geq 3$.
- (C.h) $nh^{4a} \rightarrow 0$ and $\varphi(h) \sim n^{(2/r)+b-1} / (\log(n))^2$ for large enough n and some constant $b > 0$ such that $(2/r) + b > 1/2$.

Remark 2

Note that the assumption (C. Z1) describes the structure of the subset $\mathcal{C}$ containing the functional covariates. As noted in Ferraty and Vieu (2008), this assumption is trivially satisfied in standard non-parametric problems with vector-valued covariates, but this is not necessarily true for any general abstract semi-metric space. Thus it is necessary to assume that the compact subset $\mathcal{C}$ can be written such that (C. Z1) holds. To address this issue, Ferraty et al. (2010) proposed an alternative condition based on Kolmogorov's ε -entropy (see their definition 1 in Section 2). Specifically, let us denote the Kolmogorov's ε -entropy of $\mathcal{C}$ by $\Psi_{\Psi_c}(\varepsilon)$. Ferraty et al. (2010) proposed to impose conditions on $\Psi_{\Psi_c}(\log(n)/n)$, see for example their conditions (H5) and (H6). In this paper, we will utilize the condition (C. Z1) to prove our results. However, we conjecture that it is possible to prove similar results using these alternative entropy based conditions as well. We do not take-up this problem in the present paper but it is certainly an interesting open-problem.

A.2 Proof of Result 1

We use the following two results to prove Result 1. Denote $\mathcal{Y}_i = \mathcal{Y}_i(\eta_0) = Y_i - S_i^T \eta_0$.

Lemma 1—Under regularity conditions, uniformly in $z \in \mathcal{C}$,

$$\left[\sum_{i=1}^n K_h(d(z, \mathcal{Z}_i)) \right]^{-1} \left[\sum_{i=1}^n K_h(d(z, \mathcal{Z}_i)) X_i X_i^T \right] \rightarrow E(XX^T | \mathcal{Z} = z) \text{ a.s.}$$

Lemma 2—Under regularity conditions, uniformly in $z \in \mathcal{C}$,

$$\left[\sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} \right]^{-1} \left[\sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} X_i \mathcal{Y}_i \right] = E[XX^T | \mathcal{Z} = z] \beta(z) + O(h^a) + O \left(\sqrt{\frac{\log(n)}{n\varphi(h)}} \right) \text{ a.s.}$$

Now, Using Lemma 1, Lemma 2 and positive definiteness of $E[XX^T | \mathcal{Z} = z]$, it is easy to derive that uniformly in $z \in \mathcal{C}$

$$\widehat{\beta}(z) - \beta(z) = O(h^a) + O \left(\sqrt{\frac{\log(n)}{n\varphi(h)}} \right) \text{ a.s.}$$

A.2.1 Proof of Lemma 1

Using results in Ferraty and Vieu (2004), see proof of their Lemma 3.1, we see that for $j, k = 1, \dots, p$,

$$\sup_{z \in \mathcal{C}} \left| \left[\sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} \right]^{-1} \left[\sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} X_{ij} X_{ik} \right] - E(X_j X_k | \mathcal{Z} = z) \right| = O(h^a) + O \left(\sqrt{\frac{\log(n)}{n\varphi(h)}} \right) \text{ a.s.}$$

Now, using conditions that $h \rightarrow 0$ and $\log(n)/\{n\varphi(h)\} \rightarrow 0$, we see that the right hand side is in fact $o_p(1)$ and hence the proof follows.

A.2.2 Proof of Lemma 2

Recall that under the model,

$$\mathcal{Y}_i = X_i^T \beta(\mathcal{Z}_i) + \varepsilon_i.$$

Define $\Omega(z) = E \left[\sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} \right]$. Then we have that

$$\sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} X_i \mathcal{Y}_i / \Omega(z) = A_1 + A_2,$$

where by Ferraty and Vieu (2004) uniformly in $z \in \mathcal{C}$

$$\begin{aligned} A_1 &= \sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} X_i X_i^T \beta(\mathcal{Z}_i) / \Omega(z) \\ &= E[XX^T \beta(\mathcal{Z}) | \mathcal{Z} = z] + O(h^a) + O \left(\sqrt{\frac{\log(n)}{n\varphi(h)}} \right) \text{ a.s.} \quad (3) \\ &= E[XX^T | \mathcal{Z} = z] \beta(z) + O(h^a) + O \left(\sqrt{\frac{\log(n)}{n\varphi(h)}} \right) \text{ a.s.} \end{aligned}$$

Also, using the fact that $E(\varepsilon | X, \mathcal{Z}) = 0$ we see that

$$\begin{aligned}
 A_2 &= \sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} X_i \varepsilon_i / \Omega(z) \\
 &= O(h^a) + O\left(\sqrt{\frac{\log(n)}{n\varphi(h)}}\right) a.s.
 \end{aligned}$$

Also it is easy to see (Ferraty and Vieu, 2006) that

$$\sup_{z \in C} \left| \sum_{i=1}^n K_h\{d(z, \mathcal{Z}_i)\} / \Omega(z) - 1 \right| = O\left(\sqrt{\frac{\log(n)}{n\varphi(h)}}\right) = o(1) \text{ a.s.} \quad (4)$$

Now the proof follows by combining (3) and (4).

A.3 Proof of Result 2

A.3.1 Expansion of $\hat{\eta}$

We start by noting that

$$\begin{aligned}
 0 &= n^{-1} \sum_{i=1}^n \{S_i + \widehat{\beta}_\eta^T(\mathcal{Z}_i) X_i\} \{Y_i - S_i^T \widehat{\eta} - X_i^T \widehat{\beta}(\mathcal{Z}_i, \widehat{\eta})\} \\
 &= n^{-1} \sum_{i=1}^n \{S_i + \widehat{\beta}_\eta^T(\mathcal{Z}_i) X_i\} \{Y_i - S_i^T \eta_0 - X_i^T \widehat{\beta}(\mathcal{Z}_i, \eta_0)\} - n^{-1} \sum_{i=1}^n \{S_i + \widehat{\beta}_\eta^T(\mathcal{Z}_i) X_i\} (S_i + \widehat{\beta}_\eta^T(\mathcal{Z}_i) X_i)^T (\widehat{\eta} - \eta_0).
 \end{aligned}$$

Using similar argument as in the previous section, we obtain that

$$\sup_{z \in \mathcal{C}} \left| \widehat{\beta}_{\eta,j}(z) - \beta_{\eta,j}(z) \right| = O(h^a) + O\left(\sqrt{\frac{\log n}{n\varphi(h)}}\right) a.s. \quad (5)$$

Also it is easy to see that $E(\tilde{S}X^T / \mathcal{Z}) = 0$. Now we derive

$$\begin{aligned}
 &n^{-1} \sum_{i=1}^n \{S_i + \widehat{\beta}_\eta^T(\mathcal{Z}_i) X_i\} \{Y_i - S_i^T \eta_0 - X_i^T \widehat{\beta}(\mathcal{Z}_i, \eta)\} \\
 &= n^{-1} \sum_{i=1}^n \{S_i + \widehat{\beta}_\eta^T(\mathcal{Z}_i) X_i\} \varepsilon_i \\
 &- n^{-1} \sum_{i=1}^n \{S_i + \widehat{\beta}_\eta^T(\mathcal{Z}_i) X_i\} X_i^T \{\widehat{\beta}(\mathcal{Z}_i, \eta_0) - \beta(\mathcal{Z}_i, \eta_0)\} \\
 &\quad + n^{-1} \sum_{i=1}^n \{\widehat{\beta}_\eta(\mathcal{Z}_i) - \beta_\eta(\mathcal{Z}_i)\}^T X_i \varepsilon_i \\
 &- n^{-1} \sum_{i=1}^n \{\widehat{\beta}_\eta(\mathcal{Z}_i) - \beta_\eta(\mathcal{Z}_i)\}^T X_i X_i^T \{\widehat{\beta}(\mathcal{Z}_i, \eta_0) - \beta(\mathcal{Z}_i, \eta_0)\} \\
 &= n^{-1} \sum_{i=1}^n \tilde{S}_i \varepsilon_i - A_1 + A_2 - A_3.
 \end{aligned}$$

Using (5) and the fact that $E(\varepsilon/X, \mathcal{Z}) = 0$, it is obvious that $A_2 = o_p(n^{-1/2})$. Moreover, using condition (C.h), (5) and Result 1 we have $A_3 = o_p(n^{-1/2})$. Finally, $A_1 = o_p(n^{-1/2})$ follows from Result 1 and noting that $S_i + \widehat{\beta}_\eta^T(\mathcal{Z}_i) X_i = \tilde{S}_i$ and $E(\tilde{S}X^T | \mathcal{Z}) = 0$

We observe that

$$n^{-1} \sum_{i=1}^n \{S_i + \widehat{\beta}_\eta^T(\mathcal{X}_i) X_i\} \{S_i + \widehat{\beta}_\eta^T(\mathcal{X}_i) X_i\}^T \rightarrow E(\tilde{S} \tilde{S}^T) \text{ a.s.}$$

Hence we have that

$$\widehat{\eta} - \eta_0 = \{E(\tilde{S} \tilde{S}^T)\}^{-1} n^{-1} \sum_{i=1}^n \tilde{S}_i \varepsilon_i + o_p(n^{-1/2}),$$

and the result follows.

Figure 1.

Results for the simulation study. Plotted are the actual values of $\beta_1(z)$ (1st column) and $\beta_2(z)$ (2nd column) with $z \in \mathcal{T}$ (dashed line) and the median of the estimates over all simulated data sets (solid line) along with point wise empirical 95% confidence band (dotted line). The top, middle and bottom rows correspond to three different sample sizes, $n = 200, 400$ and 1000 , respectively.

Figure 2. Results for the simulation study. Displayed are the normal Q-Q plot of estimates of η_{00} (left) and η_{01} (right) for $n = 1000$.

Table 1

Simulation results from Section 3. Presented are the scaled empirical bias (Bias = average of $n^{1/2}(\hat{\eta}_{0j} - \eta_{0j})$), empirical standard error (Emp.se = standard deviation of $n^{1/2}(\hat{\eta}_{0j} - \eta_{0j})$), estimated standard error (Est.se = average of estimated standard deviation of $n^{1/2}(\hat{\eta}_{0j} - \eta_{0j})$) and empirical coverage (at 95% nominal level) of η_{00} and η_{01} over 500 simulations.

	Parameter	Bias	Emp.se	Est.se	Emp.cov (%)
$n = 200$	η_{00}	0.031	1.43	1.22	91.6
	η_{01}	-0.043	1.34	1.18	92.0
$n = 400$	η_{00}	0.019	1.25	1.09	93.6
	η_{01}	-0.034	1.27	1.09	93.5
$n = 1000$	η_{00}	0.005	1.12	1.02	94.2
	η_{01}	-0.006	1.13	1.02	94.6

Table 2

Prediction mean squared errors for the Tecator data set for different type of varying coefficient models. Columns denoting PMSE(k) display PMSE values when the semi-metric d_k is used. In each model, the best PMSE value is given in bold face. The overall best PMSE is underlined and bold faced.

Model	PMSE(2)	PMSE(1)	PMSE(0)
$Y = \beta_0(\mathcal{Z}) + X_1\beta_1(\mathcal{Z}) + e$	2.72	5.55	8.82
$Y = \beta_0(\mathcal{Z}) + X_2\beta_2(\mathcal{Z}) + e$	1.80	1.41	2.28
$Y = \beta_0(\mathcal{Z}) + X_1\beta_1 + X_2\beta_2 + e$	1.05	0.80	2.03
$Y = \beta_0(\mathcal{Z}) + X_1\beta_1(\mathcal{Z}) + X_2\beta_2 + e$	1.29	0.88	1.58
$Y = \beta_0(\mathcal{Z}) + X_1\beta_1 + X_2\beta_2(\mathcal{Z}) + e$	1.30	<u>0.40</u>	1.47
$Y = \beta_0(\mathcal{Z}) + X_1\beta_1(\mathcal{Z}) + X_2\beta_2(\mathcal{Z}) + e$	0.99	0.56	1.48