Backfitting and smooth backfitting in varying coefficient quantile regression

YOUNG K. LEE†, ENNO MAMMEN‡ AND BYEONG U. PARK§

†Department of Statistics, Kangwon National University, Chuncheon 200-701, Korea.
E-mail: youngklee@kangwon.ac.kr

‡Department of Economics, University of Mannheim, L7, 3-5, 68131 Mannheim, Germany.
E-mail: emammen@rumms.uni-mannheim.de

§Department of Statistics, Seoul National University, Seoul 151-747, Korea.
E-mail: bupark@stats.snu.ac.kr

First version received: August 2012; final version accepted: August 2013

Summary In this paper, we study ordinary backfitting and smooth backfitting as methods of fitting varying coefficient quantile models. We do this in a unified framework that accommodates various types of varying coefficient models. Our framework also covers the additive quantile model as a special case. Under a set of weak conditions, we derive the asymptotic distributions of the backfitting estimators. We also briefly report on the results of a simulation study.

Keywords: Backfitting, Integral equation, Kernel smoothing, Quantile regression, Smooth backfitting, Varying coefficient models.

1. INTRODUCTION

There have been some recent developments for varying coefficient models that were originated by Hastie and Tibshirani (1993). Several types of varying coefficient models have been introduced as extensions of the classical linear model in a way that allows the regression coefficients to be non-parametric functions of some variable(s). A more flexible form is to take different smoothing or effect-modifying variables for different coefficient functions (i.e., to let different regression coefficients be functions of different variables). This type is a structural regression model, which is known to be an effective way of avoiding the curse of dimensionality when there are many covariates. For this type, most works have been focused on the mean regression context. Some recent works on this type include Yang et al. (2006), Roca-Pardinas and Sperlich (2010) and Lee et al. (2012a). Park et al. (2013) have provided a comprehensive review of several different types of varying coefficient models along with various statistical problems and methods associated with those models. Lee et al. (2012b) have proposed and studied a unifying approach to all types of varying coefficient models with powerful techniques for fitting the fully flexible model.

Varying coefficient models have many economics applications. In a wage model in labour economics, for example, it might be desirable to allow marginal returns to education to vary with the level of schooling and with work experience. For other applications, see Hong and
Lee (2003), Li et al. (2002) and Cai et al. (2006), for example. In these applications, however, quantile regression is a more natural or more informative approach than mean regression. Quantile regression gives a picture of the entire conditional distributions, and can be used to build conditional prediction intervals. Also, it provides a useful tool for verifying the presence of conditional heteroscedasticity; see, e.g., Furno (2004). There have been only a few works for varying coefficient quantile regression. Among these, Kim (2007) and Wang et al. (2009) considered sieve estimation, and Honda (2004) worked on kernel estimation. In these works, the single smoothing variable case was considered (i.e., the case where all coefficients are functions of the same smoothing variable). In kernel estimation of models of this type, the standard technique applies and gives directly the proper estimators of coefficient functions. Thus, there is no need for backfitting.

In this paper, we discuss quantile regression for the fully flexible varying coefficient model, which was introduced and studied by Lee et al. (2012b) in the mean regression context. We consider the applications of the ordinary backfitting (BF) approach and the smooth backfitting approach (SBF). The two procedures were proposed by Buja et al. (1989) and Mammen et al. (1999), respectively, in the additive mean regression setting. In that setting, the theory for BF was developed by Opsomer and Ruppert (1997), and the SBF approach was further studied in related problems; see, e.g., Mammen and Park (2005), Mammen and Park (2006) and Yu et al. (2008). The marginal integration technique of Linton and Nielsen (1995) as another way of fitting additive models is not considered here, because it is widely accepted that it suffers from the curse of dimensionality. We provide backfitting algorithms for BF and SBF and discuss their statistical properties in the quantile regression setting. In particular, we show that both BF methods produce rate-optimal estimators and derive their asymptotic distributions.

For covariates \( X_1, \ldots, X_D \), we write \( X = (X_1, \ldots, X_D)\top \) and we assume that the quantile function \( m(x) \) satisfies

\[
m(x) = x_1 \left( \sum_{k \in I_1} f_{1k}(x_k) \right) + \cdots + x_d \left( \sum_{k \in I_d} f_{dk}(x_k) \right),
\]

where the index sets \( I_j \subset \{1, 2, \ldots, D\} \) are known and may not be disjoint. Without loss of generality, we assume that each \( I_j \) does not include \( j \). The reason for this is that we allow \( X_j \equiv 1 \) for some \( 1 \leq l \leq d \), and for such \( X_l \) we may put \( X_j f_{jl}(X_j) \) into \( X_l \sum_{k \in I_l} f_{lk}(X_k) \) in the form of \( X_l g(X_j) \), where \( g(X_j) = X_j f_{jl}(X_j) \). Here, \( d \leq D \). The level \( \alpha \) of the quantile is fixed throughout the paper:

\[
P(Y \leq m(x)|X = x) = \alpha.
\]

The covariates that enter into one of the coefficient functions \( f_{jk} \) (i.e., \( X_k \) for \( k \in C \equiv \cup_{j=1}^d I_j \)) are of continuous type. For simplicity, we assume that \( X_k \) with \( k \in C \) are supported on the interval \([0, 1]\). We allow some of the covariates \( X_k \), for \( 1 \leq j \leq d \), to be discrete random variables. In particular, the model may include a variable \( X_j \equiv 1 \) for one \( j \in \{1, 2, \ldots, d\} \). We also allow that \( C \) and \( \{1, 2, \ldots, d\} \) may have common indices. Let \( C_0 = C \cap \{1, 2, \ldots, d\} \).

The form of the model (1.1) unifies various types of varying coefficient models. First, it reduces to the single smoothing variable case studied by Kim (2007), Wang et al. (2009) and Honda (2004), if one takes \( D = d + 1 \) and \( I_j = \{d + 1\} \) for all \( 1 \leq j \leq d \). It also covers the varying coefficient regression models of Xue and Yang (2006) and Lee et al. (2012a), where \( \{X_1, \ldots, X_d\} \) and the group of smoothing variables \( \{X_k : k \in C\} \) do not have any variable in common (i.e., \( C_0 \) is empty). Furthermore, the framework includes even the additive quantile
model studied by Horowitz and Lee (2005) and Lee et al. (2010) as a special case with the choice \( d = 1 \) and \( X_1 \equiv 1 \).

The functions \( f_{jk} \) in the representation (1.1) are not identifiable. To make all \( f_{jk} \) identifiable, we put the following constraints for non-negative weight functions \( w_k \):

\[
\int f_{jk}(x_k) w_k(x_k) \, dx_k = 0, \quad k \in C, \ 1 \leq j \leq d; \\
\int x_k f_{jk}(x_k) w_k(x_k) \, dx_k = 0, \quad j, k \in C_0.
\]

The first constraint of (1.2) is typical in additive models and it is for not allowing the two different representations \( f_{j1}(x_1) + f_{j2}(x_2) = g_{j1}(x_1) + g_{j2}(x_2) \), for example, where \( g_{j1} = f_{j1} + c \) and \( g_{j2} = f_{j2} - c \) for some constant \( c \). The second constraint is also required. It is to avoid the two different representations \( x_1 f_{12}(x_2) + x_2 f_{21}(x_1) = x_1 g_{12}(x_2) + x_2 g_{21}(x_1) \), for example, where \( g_{12}(x_2) = f_{12}(x_2) + x_2 \) and \( g_{21}(x_1) = f_{21}(x_1) - x_1 \). With these constraints, all \( f_{jk} \) are identifiable; see Lee et al. (2012b) for a proof. We can also rewrite the model (1.1) as

\[
m(x) = \sum_{j=1}^{d} a_j x_j + \sum_{j,k \in C_0} a_{jk} x_j x_k + \sum_{j=1}^{d} x_j \left( \sum_{k \in I_j} f_{jk}(x_k) \right),
\]

which \( f_{jk} \) satisfy (1.2).

In the description of our methods and in our theory, we also make use of a different representation of the model (1.3). In this representation of the model, we collect those coefficients that are functions of the same continuous covariate and put them together as an additive component. Suppose that, among \( X_1, \ldots, X_d \) in the model (1.3), there are \( r \ (0 \leq r \leq d) \) variables whose indices do not enter into \( C \). Without loss of generality, we denote these by \( X_1, \ldots, X_r \). Let \( p = D - r \geq 2 \) be the number of indices in \( C \). Thus, \( C = \{ r + 1, \ldots, r + p \} \) and \( C_0 = \{ r + 1, \ldots, d \} \). The case \( p = 1 \) is not considered in this paper because, in the latter case, no backfitting is needed. Define

\[
\tilde{X}_k = \{ X_j : r + k \in I_j, \ 1 \leq j \leq d \}, \quad 1 \leq k \leq p.
\]

The vector \( \tilde{X}_k \) is the collection of all \( X_j \), for \( 1 \leq j \leq d \), that have interactions with \( X_{r+k} \) in the form of \( X_j f_{j,r+k}(X_{r+k}) \). Thus, \( \tilde{X}_k \) does not include \( X_{r+k} \) as its element because the index set \( I_j \) does not contain \( j \). Let \( d_k \) denote the number of the index sets \( I_j \) that contain \( r + k \). Thus, \( \tilde{X}_k \) is of \( d_k \)-dimension. Likewise, for a given vector \( x \), we denote the above rearrangements of \( x \) by \( \tilde{x}_k \), \( 1 \leq k \leq p \). Also, define \( f_k = \{ f_{j,r+k} : r + k \in I_j, \ 1 \leq j \leq d \} \) for \( 1 \leq k \leq p \). Then, the model (1.3) can be represented as

\[
m(x) = \sum_{j=1}^{d} a_j x_j + \sum_{j,k \in C_0} a_{jk} x_j x_k + \tilde{x}_1^\top f_1(x_{r+1}) + \cdots + \tilde{x}_p^\top f_p(x_{r+p}).
\]

In the next section, we describe the BF and SBF algorithms to fit the model (1.5) and discuss their theoretical properties. Section 3 is devoted to the proof of the main theorem in Section 2. The numerical properties of the backfitting estimators are presented in Section 4.
2. BACKFITTING METHODS

We introduce two kernel estimators, one based on the ordinary BF and the other based on the SBF approach. Model (1.1) can be rewritten as (1.3) with constraints (1.2). Both the ordinary BF and the SBF estimator require starting values $\hat{f}_{BF,0}^\ell$ and $\hat{f}_{SBF,0}^\ell$ ($1 \leq \ell \leq p$), respectively. Furthermore, we need initial estimators $\hat{a}_j^{I,0}$ and $\hat{a}_{jk}^{I,0}$, for $l \in \{BF, SBF\}$, of $a_j$ ($1 \leq j \leq d$) and $a_{jk}$ ($j < k; j, k \in C_0$). We assume that the initial estimators $\hat{f}_{BF,0}^\ell$ and $\hat{f}_{SBF,0}^\ell$ fulfil the constraints (1.2), but we define their updates $\hat{f}_{BF,1}^\ell$ and $\hat{f}_{SBF,1}^\ell$ ($k \geq 1$), respectively, which might not fulfill (1.2). We do not update the estimators of $a_j$ and $a_{jk}$; see (2.1) and (2.2). After stopping the iteration of a BF algorithm after $k^*$ steps, say, it is easy to adjust the last updated estimators, $\hat{f}_{BF,k^*}^\ell$ or $\hat{f}_{SBF,k^*}^\ell$, by subtracting constants or linear functions so that (1.2) is fulfilled for the normalized estimators. We refer to (4.7) in Lee et al. (2012b) for the normalizing procedure. We denote these estimators by $\hat{f}_{BF}^\ell$ and $\hat{f}_{SBF}^\ell$, respectively. We discuss the choice of $k^*$ below. The final estimators $\hat{a}_j^l$ and $\hat{a}_{jk}^l$ of $a_j$ and $a_{jk}$ are defined by

$$
\sum_{j=1}^d \hat{a}_j^l x_j + \sum_{j<k, j \leq l} \hat{a}_{jk} x_j x_k + \sum_{\ell=1}^p \tilde{x}_\ell^l \hat{f}_\ell^l (x_{r+\ell})
= \sum_{j=1}^d \hat{a}_j^{I,0} x_j + \sum_{j<k, j \leq l} \hat{a}_{jk}^{I,0} x_j x_k + \sum_{\ell=1}^p \tilde{x}_\ell^l \hat{f}_{I,0}^l (x_{r+\ell}), \quad l \in \{BF, SBF\}.
$$

For theoretical purposes, we make some assumptions on the starting functions, $\hat{f}_{BF,0}^\ell$ and $\hat{f}_{SBF,0}^\ell$; see Assumptions 3.5 and 3.6 in Section 3. Some choices that fulfill these conditions are discussed in Section 2 of Lee et al. (2010) in the setting of the additive model. Similar choices can be made for our varying coefficient model. In their Section 2, Lee et al. (2010) also discussed the possibility of relaxing the conditions; see also a remark given immediately after Theorem 3.1 in Section 3. For given starting functions $\hat{f}_{BF,0}^\ell$ and $\hat{f}_{SBF,0}^\ell$, the initial estimators $\hat{a}_j^{I,0}$ and $\hat{a}_{jk}^{I,0}$ for $l \in \{BF, SBF\}$ can be obtained by minimizing

$$
\sum_{i=1}^n \int \tau_\alpha (Y^i - \sum_{j=1}^d \alpha_j^{I,0} X^i_j - \sum_{j<k, j \leq l} \alpha_{jk}^{I,0} X^i_j X^i_k - \sum_{\ell=1}^p \tilde{x}_\ell^l \hat{f}_{I,0}^l (X^i_{r+\ell}))
$$

with respect to $\alpha_j^{I,0}$ and $\alpha_{jk}^{I,0}$, where $\tau_\alpha$ denotes the so-called check function defined by $\tau_\alpha (u) = u \{\alpha - I(u < 0)\}$. In our simulation study in Section 4, we have taken $\hat{f}_{BF,0}^\ell = \hat{f}_{SBF,0}^\ell = 0$ for all $1 \leq \ell \leq p$ and we have found that it works very well.

The ordinary BF algorithm runs in iterative cycles. Let $K_g$ be a kernel function with bandwidth $g$; see Assumption 3.2. Then, in the $j$th step of the $k$th cycle, the estimator $\hat{f}_{BF,k}^j$ is updated as follows:

$$
\hat{f}_{BF,k+1}^j (u) = \arg \min_{\theta \in \Theta_j} \sum_{i=1}^n \tau_\alpha (Y^i - \sum_{j=1}^d \hat{a}_j^{BF,0} X^i_j - \sum_{j<k, j \leq l} \hat{a}_{jk}^{BF,0} X^i_j X^i_k - \tilde{x}_\ell^l \hat{f}_{I,0}^l (X^i_{r+\ell}))
- \sum_{\ell=1}^p \tilde{x}_\ell^l \hat{f}_{BF,k}^j (X^i_{r+\ell}) - \sum_{\ell=j+1}^p \tilde{x}_\ell^l \hat{f}_{BF,k}^j (X^i_{r+\ell}) K_h_j (u, X^i_{r+j}). \quad (2.1)
$$
To simplify the mathematical arguments, the minimization in (2.1) is undertaken over a compact set $\Theta_j \subset \mathbb{R}^{d_j}$. It is assumed that all values of the function $f_j$ lie in the interior of $\Theta_j$. As in the case of mean regression, the ordinary BF estimator is not defined as a solution of a global minimization problem.

The SBF estimator is also computed by an iterative scheme. The estimate of $f_j$ is updated by

$$
\hat{f}_{j,[k+1]}(u) = \arg\min_{\theta \in \Theta_j} \sum_{i=1}^{n} \tau_{a}(Y^{i} - \sum_{j=1}^{d} \hat{a}_{j}^{SBF,[0]}X^{i}_{j} - \sum_{j<k, j \neq k \in \theta_{0}} \hat{a}_{j}^{SBF,[0]}X^{i}_{j}X^{i}_{k} - \tilde{X}_{j}^{\top} \theta
$$

$$
- \sum_{\ell=1}^{j-1} \tilde{X}_{\ell}^{\top} \hat{t}_{\ell}^{SBF,[k+1]}(z_{\ell}) - \sum_{\ell=1}^{p} \tilde{X}_{\ell}^{\top} \hat{t}_{\ell}^{SBF,[k]}(z_{\ell})
$$

$$
\times \prod_{\ell=1, \neq j}^{p} K_{h_{l}}(z_{\ell}, X^{i}_{r+\ell}) d z_{\ell} : K_{h_{j}}(u, X^{i}_{r+j}).
$$

(2.2)

Here, the integration runs over the support of $(X^{i}_{r+1}, \ldots, X^{i}_{r+j-1}, X^{i}_{r+j+1}, \ldots, X^{i}_{r+p})$. For $k \to \infty$, this is an iterative scheme for obtaining $\hat{f}_{j}^{SBF}$, $1 \leq j \leq p$, which minimize

$$
\sum_{i=1}^{n} \tau_{a}(Y^{i} - \sum_{j=1}^{d} \hat{a}_{j}^{SBF,[0]}X^{i}_{j} - \sum_{j<k, j \neq k \in \theta_{0}} \hat{a}_{j}^{SBF,[0]}X^{i}_{j}X^{i}_{k} - \tilde{X}_{j}^{\top} \hat{t}_{\ell}^{SBF}(z_{\ell}))^{2}
$$

$$
\times \prod_{\ell=1}^{p} K_{h_{l}}(z_{\ell}, X^{i}_{r+\ell}) d z_{\ell},
$$

(2.3)

where the integration is over the support of $(X^{i}_{r+1}, \ldots, X^{i}_{r+p})$.

As mentioned above, we assume for theoretical purposes that the iterative algorithm stops after a finite number $k^{*}$ of cycles; see Theorem 3.1 and the discussion that follows for a theoretical choice of $k^{*}$. A stopping rule in practice would be a value $k^{*}$ for which the difference between the two updates at the $(k^{*} - 1)$th cycle and the $k^{*}$th cycle is sufficiently small. One measure for the difference is

$$
\sum_{j=1}^{p} \int (\hat{f}_{j}^{[k^{*}]}(u) - \hat{f}_{j}^{[k^{*}-1]}(u))^{2} du, \quad l \in \{BF, SBF\}.
$$

In this case, we should use the normalized versions of the updates that satisfy the constraints (1.2).

Comparing the two BF equations (2.1) and (2.2), we see that the SBF estimators require multiple integration while the BF estimators do not. The dimension of the multiple integration for SBF is $p - 1$ so that the computational costs increase as $p$ becomes high. This is a drawback of the SBF method in quantile regression. We note that this is not the case in the mean regression setting where we only need a single-dimensional integration regardless of the dimension $p$. It is possible to speed up the computing time for SBF by applying a well-devised Monte Carlo method for the numerical integration. For a discussion on this issue in the setting of quasi-likelihood additive regression, we refer to Section 5 of Yu et al. (2008). The SBF method, however, gives
more stable estimators, as we have found in our simulation study in Section 4. The method is found to produce estimators with better mean integrated squared error performance than the BF method in all settings of our simulation study.

3. ASYMPTOTIC THEORY

We develop a complete asymptotic theory for these estimators. We do this for the case where i.i.d. observations \((X^i, Y^i)\) are made on the random vector \((X, Y)\). To keep the presentation simple, we assume that \(a_j \equiv 0\) and \(a_{jk} \equiv 0\), and we also take \(\hat{a}^{[0]}_j \equiv 0\) and \(\hat{a}^{[0]}_{jk} \equiv 0\) for \(i \in \{\text{BF, SBF}\}\). The theory we present here remains valid in general cases because \(a_j\) and \(a_{jk}\) can be estimated at a faster rate than \(f_{jk}\).

To discuss the theoretical properties of the BF estimators, we make the following assumptions.

**Assumption 3.1.** It holds that the product measure \(\prod_{j=1}^{D} P_{X_j}\) has a density with respect to the distribution \(P_X\) of \(X\) that is bounded away from zero and infinity on the support of \(P_X\). Here, \(P_{X_j}\) is the marginal distribution of \(X_j\). The marginal distributions are absolutely continuous with respect to Lebesgue measure, or they are discrete measures with a finite support. Furthermore, the weight functions \(w_j\) for \(j \in \mathcal{C}\) in the constraints (1.2) are chosen so that \(w_j/p_{X_j}\) is bounded away from zero and infinity on the support of \(P_{X_j}\). Here, \(P_{X_j}\) is the density of \(X_j\). The smallest eigenvalues of the matrices \(E[\hat{X}_j\hat{X}_j^\top|X_{r+j} = z_j]\) for \(1 \leq j \leq p\) are bounded away from zero for \(z_j\) in the support of \(P_{X_{r+j}}\). The densities \(p_{X_j}\) for \(r + 1 \leq j \leq r + p\) are bounded away from zero on \([0, 1]\). The weight functions \(w_j\) are continuously differentiable, fulfil \(w_j(0) = w_j(1) = 0\), \(w_j(x_j) \geq 0\) for \(x_j \in [0, 1]\) and \(\int w_j(x_j) dx_j > 0\) for \(j \in \mathcal{C}\).

Assumption 3.1 is a modification of Assumption A0 in Lee et al. (2012b). It guarantees that for each function \(m\) there exists only one tuple \((f_1, \ldots, f_p)\) with (1.5) subject to the constraint (1.2).

We now make assumptions on the kernels and the order of the bandwidths. We use boundary corrected kernels. Assumption 3.2, in particular, concerns the shape of the kernel at the boundary. Note that we need different kinds of boundary correction for the ordinary BF and for the SBF estimator.

**Assumption 3.2.** There exist constants \(c_K\), \(C_K\), \(C_S\), \(C'_S > 0\) such that: (a) for all \(u \in [0, 1]\), the kernel \(K_s(u, \cdot)\) is positive, bounded by \(C_K g^{-1}\), has bounded support in \([u - C_S g, u + C_S g]\) and is Lipschitz continuous with Lipschitz constant bounded by \(C_K g^{-2}\); (b) the kernel \(K_s(u, \cdot)\) has a second derivative with respect to \(u\) that is bounded by \(C_D g^{-3}\), and fulfils \(\int K_s(u, v) dv \geq c_K\); (c) for \(C'_S g \leq u, v \leq 1 - C'_S g\), it holds that \(K_s(u, v) = g^{-1} K(g^{-1}(u - v))\) for a function \(K\) with \(\int K(v) dv = 1\) and \(\int v K(v) dv = 0\); (d) in the case of the ordinary BF, the kernel satisfies \(\int K_s(u, u) dv = 1\) and \(\int K_s(u, v)(u - v) dv = 0\), while in the case of the SBF \(\int K_s(u, v) du = 1\).

**Assumption 3.3.** The bandwidths \(h_1, \ldots, h_p\) are of order \(n^{-1/5}\) with \(n^{1/5} h_j \to c_j\) as \(n \to \infty\) for some constants \(0 < c_j < \infty\).

**Assumption 3.4.** The conditional density \(p_{\epsilon|x}(0|x)\) of \(\epsilon \equiv Y - m(X)\) given \(X = x\) is bounded away from zero and infinity on the support of \(P_X\). Furthermore, it satisfies the following uniform
\[ |p_{\xi}|(e|\mathbf{x}) - p_{\xi}|(0|\mathbf{x})| \leq C_1|e|, \]

for \( \mathbf{x} \) in the support of \( P_X \) and for \( e \) in a neighbourhood of 0, with a constant \( C_1 > 0 \) that does not depend on \( \mathbf{x} \). The partial derivative \( \partial p_X(\mathbf{x})/\partial \mathbf{x}^\top \) of the joint density function \( p_X \) exists and is continuous in \( \mathbf{x}' \equiv (x_{r+1}, \ldots, x_{r+p}) \) for all \( (x_1, \ldots, x_r) \). The components of \( \hat{f}_j \) are twice continuously differentiable.

Assumption 3.4 is a standard smoothing condition. We now make some assumptions on the starting values \( \hat{f}_{\ell}^{BF,[0]} \) and \( \hat{f}_{\ell}^{SBF,[0]} \) of the BF algorithms. For some \( 0 < \rho \leq 1, \delta_1, \delta_2 > 0 \) and \( \xi \geq 0 \) we assume the following.

**Assumption 3.5.** For \( j = 1, \ldots, p \), it holds for \( i = BF \) and \( i = SBF \) that
\[
\sup_{C\delta_j \leq u \leq 1 - C\delta_j} |\hat{f}_j^{i,[0]}(u) - f_j(u)| = O_p(n^{-(2/5)+\delta_1}),
\]
\[
\sup_{0 \leq u \leq 1} |\hat{f}_j^{i,[0]}(u) - f_j(u)| = O_p(n^{-(1/5)+\delta_1}).
\]

**Assumption 3.6.** There exist random functions \( g_1, \ldots, g_p \) with derivatives that fulfil the Lipschitz condition,
\[
|g_j'(u) - g_j'(v)| \leq C|u - v|^\rho n^\xi,
\]
for \( j = 1, \ldots, p \) and \( u, v \in [0, 1] \). Furthermore, these functions satisfy
\[
\sup_{C\delta_j \leq u \leq 1 - C\delta_j} |\hat{f}_j^{i,[0]}(u) - g_j(u)| = O_p(n^{-(2/5)-\delta_2}),
\]
\[
\sup_{0 \leq u \leq 1} |\hat{f}_j^{i,[0]}(u) - g_j(u)| = O_p(n^{-(1/5)+\delta_2}),
\]
for \( i = BF \) and \( i = SBF \).

For our final assumption, we need some additional notation. Put
\[
W_{jk}(z_j, z_k) = E[p_{\xi}|X(0)|\tilde{X}_j\tilde{X}_k^\top|X_{r+j} = z_j, X_{r+k} = z_k]p_{j,k}(z_j, z_k),
\]
\[
W_{j}(z_j) = E[p_{\xi}|X(0)|\tilde{X}_j\tilde{X}_j^\top|X_{r+j} = z_j]p_j(z_j),
\]
\[
W_{j}^s(z_j) = E[\tilde{X}_j\tilde{X}_j^\top|X_{r+j} = z_j]p_{j}(z_j),
\]
and
\[
\Sigma_j(z_j) = \frac{\alpha(1 - \alpha)}{\epsilon_j} \int K^2(t) dt W_{j}(z_j)^{-1}W_{j}^s(z_j)W_{j}(z_j)^{-1}.
\]

Here, \( p_j \) denotes the density of \( X_{r+j} \) and \( p_{j,k} \) denotes the density of \( (X_{r+j}, X_{r+k}) \). Note that Assumptions 3.1 and 3.4 imply that the smallest eigenvalues of the matrices \( W_{j}(z_j) \), for \( 1 \leq j \leq p \), are bounded away from zero for \( z_j \in [0, 1] \).

**Assumption 3.7.** The elements of the matrices \( W_{j}(z_j) \) and \( W_{jk}(z_j, z_k) \) have second-order derivatives with respect to \( z_j \) that are bounded over \( z_j \in [0, 1] \) and \( z_k \in [0, 1] \), \( 1 \leq j \neq k \leq p \).
Furthermore, the elements of $W_{jk}(z_j, z_k)$ and their first-order derivatives with respect to $z_j$ are continuous in $z_k$.

We now introduce notation for the asymptotic biases and variances of the BF estimators. Define $p_{j,X}^{(1)}(x) = \partial p_x(x)/\partial x_j$, and $p_j^{(1)}(0|x) = \partial p_x(0|x)/\partial x_j$, and denote the first and second derivatives of $f_j^*$ by $f_j^*$ and $f_j''$, respectively. Put

$$
\tilde{\beta}_j^*(z_j) = W_{j,k}(z_j)^{-1}c_j^2 E \left[ \tilde{X}_j \tilde{X}_k^\top f_k'(X_{r+k}) \left( p_{k,X}(0 | X) \frac{p_{k,X}(X)}{p_X(X)} + p_{k}_{\ell}(0 | X) \right) \right]
$$

$$
+ \Delta_{jk} \tilde{X}_k^\top f_k'(X_{r+k}) p_{\ell|X}(0 | X) X_{r+j} = z_j \right] p_j(z_j) \int t^2 K(t) dt,
$$

for $1 \leq j, k \leq p$. Here, $\Delta_{jk}$ is a vector that has the same dimension as $\tilde{X}_j$ and has elements $\Delta_{jk,\ell}$ such that $\Delta_{jk,\ell} = 1$ if, in our rearrangement (1.4), the $\ell$th element of $\tilde{x}_j$ equals $x_{r+j}$, and $\Delta_{jk,\ell} = 0$ otherwise. Thus, if $\tilde{x}_j$ does not contain $x_{r+k}$ as its element, then $\Delta_{jk}$ is a zero vector. For the ordinary BF estimator, let $\beta_j^{*,BF}(z) = (\tilde{\beta}_j^{*,BF}(z_j) : 1 \leq j \leq p)$ be the solution of the following system of integral equations,

$$
\beta_j^{*,BF}(z_j) = \tilde{\beta}_j^*(z_j) + \frac{1}{2} c_j^2 f_j''(z_j) \int \int t^2 K(t) dt
$$

$$
- \sum_{k \neq j} (W_{j,k}(z_j))^{-1} W_{jk}(z_j, z_k) \beta_k^{*,BF}(z_k) dz_k, \quad 1 \leq j \leq p,
$$

and put $\beta_j^{BF}(z_j)$ to be the normalized versions of $\beta_j^{*,BF}(z_j)$ so that they satisfy the constraints (1.2). For the SBF estimator, define $\beta_j^{*,SBF}(z) = (\tilde{\beta}_j^{*,SBF}(z_j) : 1 \leq j \leq p)$ be the solution of

$$
\beta_j^{*,SBF}(z_j) = \sum_{k=1}^p \tilde{\beta}_j^*(z_j) - \sum_{k \neq j} (W_{j,k}(z_j))^{-1} W_{jk}(z_j, z_k) \beta_k^{*,SBF}(z_k) dz_k, \quad 1 \leq j \leq p,
$$

and put $\beta_j^{SBF}(z_j)$ to be the normalized versions of

$$
\beta_j^{*,SBF}(z_j) + \frac{1}{2} c_j^2 f_j''(z_j) \int \int t^2 K(t) dt,
$$

so that they satisfy the constraints (1.2).

**Theorem 3.1.** Assume that Assumptions 3.1–3.7 hold with $\xi \geq 0$, $\delta_1, \delta_2 > 0$ small enough. Let $p_X$ denote the density of $X^c \equiv (X_{r+1}, \ldots, X_{r+p})$. Then, we obtain for $\hat{f}_j^i = \hat{f}_j^{i,[k^*]}$ (i = BF and k = SBF) with $k^* = C_{iter,i} \log n$ for an appropriate choice of $C_{iter,i}$ that for all $z$ in the interior of the support of $p_X$, $n^{2/5}(\hat{f}_j(z_j) - f_j(z_j))$ are jointly asymptotically normal with mean $(\beta_1(z_1)^\top, \ldots, \beta_p(z_p)^\top)^\top$ and variance diag($\Sigma_j(z_j)$).

We conjecture that the ordinary BF and the SBF also work for less accurate starting values $\hat{f}_j^{[0]}$ and also for the limits of the iterative BF algorithms ($k^* = \infty$), and that the same asymptotic limit applies as in Theorem 3.1. For two reasons, we need in our proof that $\hat{f}_j^{[0]}$ converges with a certain rate to the true function, and that $k^*$ is finite. First, in the proof, we approximate the non-smooth and non-linear criteria given at (2.1) and (2.2) by smooth linear approximations.
Table 1. MISE, ISB and IV of the BF and SBF estimators: $X_j$ uncorrelated.

| Sample size | Coefficient function | Measure | $\alpha = 0.2$ | $\alpha = 0.5$ | $\alpha = 0.8$ |
|-------------|----------------------|---------|--------------|--------------|--------------|
| $n = 400$   | $f_{13}$             | MISE    | 0.0470       | 0.0460       | 0.0223       | 0.0219       | 0.0518       | 0.0507       |
|             |                      | ISB     | 0.0385       | 0.0386       | 0.0177       | 0.0175       | 0.0446       | 0.0442       |
|             |                      | IV      | 0.0085       | 0.0074       | 0.0046       | 0.0044       | 0.0072       | 0.0065       |
| $f_{23}$    | MISE                 | 0.0076   | 0.0065       | 0.0050       | 0.0044       | 0.0077       | 0.0065       |
|             |                      | ISB     | 0.0010       | 0.0012       | 0.0012       | 0.0013       | 0.0012       | 0.0012       |
|             |                      | IV      | 0.0066       | 0.0053       | 0.0037       | 0.0031       | 0.0065       | 0.0053       |
| $f_{24}$    | MISE                 | 0.0236   | 0.0211       | 0.0146       | 0.0133       | 0.0248       | 0.0229       |
|             |                      | ISB     | 0.0045       | 0.0048       | 0.0048       | 0.0047       | 0.0072       | 0.0068       |
|             |                      | IV      | 0.0191       | 0.0163       | 0.0098       | 0.0086       | 0.0177       | 0.0161       |
| $n = 1000$  | $f_{13}$             | MISE    | 0.0317       | 0.0316       | 0.0093       | 0.0091       | 0.0359       | 0.0354       |
|             |                      | ISB     | 0.0264       | 0.0266       | 0.0064       | 0.0064       | 0.0311       | 0.0312       |
|             |                      | IV      | 0.0053       | 0.0049       | 0.0029       | 0.0027       | 0.0048       | 0.0043       |
| $f_{23}$    | MISE                 | 0.0051   | 0.0042       | 0.0026       | 0.0024       | 0.0049       | 0.0039       |
|             |                      | ISB     | 0.0004       | 0.0004       | 0.0003       | 0.0003       | 0.0004       | 0.0003       |
|             |                      | IV      | 0.0048       | 0.0038       | 0.0023       | 0.0021       | 0.0045       | 0.0036       |
| $f_{24}$    | MISE                 | 0.0112   | 0.0103       | 0.0058       | 0.0053       | 0.0109       | 0.0096       |
|             |                      | ISB     | 0.0009       | 0.0010       | 0.0010       | 0.0010       | 0.0011       | 0.0010       |
|             |                      | IV      | 0.0103       | 0.0094       | 0.0048       | 0.0043       | 0.0098       | 0.0086       |

At that point, we need the initial functions to be not far away from the true functions. Second, for the BF method in additive mean regression, we can show that the BF algorithm converges with a geometric speed. The main argument in the background for this claim is to prove that the updating operator is a contraction. Updating algorithms for the minimizers of the smooth linear approximations mentioned above also have a geometric speed of convergence, and the limits of the algorithms have the asymptotic distributions given in Theorem 3.1. However, the smooth linear approximations of the original non-smooth and non-linear criteria produce errors in the resulting estimators and these approximation errors increase as the iteration goes on. For this reason, we need to stop after a finite number of iterations in order to control the errors so that they remain negligible in comparison with the first-order estimation error, which is $n^{-2/5}$. As can be seen from the results in Theorem 3.1, we allow the number of iterations to increase with the sample size and we choose an appropriate speed that gives the same asymptotic distributions as the limits of the approximating smooth linear BF algorithms.

4. NUMERICAL RESULTS

In this section, we briefly report the simulation results for BF and SBF. In the simulation, we have considered the following model:

$$Y^i = 4X_1^i(X_3^i - 0.5)^2 + X_2^i(\exp(X_3^i - 1) + \cos(2\pi X_4^i)) + (X_1^i \exp(-X_3^i) + X_2^i/2)U^i.$$
Table 2. MISE, ISB and IV of BF and SBF estimators: $X_j$ correlated.

| Sample size | Coefficient function | Measure | $\alpha = 0.2$ | $\alpha = 0.5$ | $\alpha = 0.8$ |
|-------------|----------------------|---------|----------------|----------------|----------------|
| $n = 400$   | $f_{13}$             | MISE    | 0.0421         | 0.0406         | 0.0235         |
|             |                      |         | 0.0226         | 0.0625         | 0.0610         |
|             |                      | ISB     | 0.0339         | 0.0333         | 0.0181         |
|             |                      |         | 0.0179         | 0.0523         | 0.0515         |
|             |                      | IV      | 0.0082         | 0.0073         | 0.0054         |
|             |                      |         | 0.0047         | 0.0101         | 0.0095         |
|             | $f_{23}$             | MISE    | 0.0089         | 0.0073         | 0.0048         |
|             |                      |         | 0.0046         | 0.0082         | 0.0073         |
|             |                      | ISB     | 0.0014         | 0.0015         | 0.0012         |
|             |                      |         | 0.0015         | 0.0010         | 0.0014         |
|             |                      | IV      | 0.0075         | 0.0058         | 0.0036         |
|             |                      |         | 0.0031         | 0.0073         | 0.0058         |
|             | $f_{24}$             | MISE    | 0.0271         | 0.0242         | 0.0168         |
|             |                      |         | 0.0154         | 0.0260         | 0.0233         |
|             |                      | ISB     | 0.0067         | 0.0064         | 0.0058         |
|             |                      |         | 0.0057         | 0.0065         | 0.0064         |
|             |                      | IV      | 0.0204         | 0.0179         | 0.0110         |
|             |                      |         | 0.0097         | 0.0195         | 0.0169         |
| $n = 1000$  | $f_{13}$             | MISE    | 0.0309         | 0.0301         | 0.0109         |
|             |                      |         | 0.0105         | 0.0419         | 0.0408         |
|             |                      | ISB     | 0.0248         | 0.0245         | 0.0071         |
|             |                      |         | 0.0070         | 0.0345         | 0.0341         |
|             |                      | IV      | 0.0061         | 0.0057         | 0.0038         |
|             |                      |         | 0.0035         | 0.0075         | 0.0067         |
|             | $f_{23}$             | MISE    | 0.0046         | 0.0038         | 0.0028         |
|             |                      |         | 0.0026         | 0.0051         | 0.0046         |
|             |                      | ISB     | 0.0003         | 0.0004         | 0.0005         |
|             |                      |         | 0.0005         | 0.0005         | 0.0006         |
|             |                      | IV      | 0.0043         | 0.0034         | 0.0023         |
|             |                      |         | 0.0021         | 0.0046         | 0.0040         |
|             | $f_{24}$             | MISE    | 0.0119         | 0.0107         | 0.0065         |
|             |                      |         | 0.0058         | 0.0122         | 0.0112         |
|             |                      | ISB     | 0.0010         | 0.0009         | 0.0009         |
|             |                      |         | 0.0008         | 0.0008         | 0.0008         |
|             |                      | IV      | 0.0109         | 0.0097         | 0.0057         |
|             |                      |         | 0.0050         | 0.0114         | 0.0104         |

There are two scenarios for the choice of covariates $X_j$: (1) $(X_1^i, X_2^i)$ were i.i.d. from $N(0, I)$ where $I$ was the $2 \times 2$ identity matrix, and $(X_3^i, X_4^i)$ independent of $(X_1^i, X_2^i)$ were i.i.d. from $N(1/2, I)$ truncated on $[0, 1]^2$ where $I = (1, 1)$; (2) $(X_1^i, X_2^i)$ were i.i.d. from $N(0, J_1)$ where $J_1$ was the $2 \times 2$ matrix $(1; 0.7; 1)$ so that the correlation between $X_1^i$ and $X_2^i$ was 0.7, and $(X_3^i, X_4^i)$ independent of $(X_1^i, X_2^i)$ were i.i.d. from $N(1/2, J_2)$ truncated on $[0, 1]^2$ where $J_2 = (1; 0.9; 1)$. The error terms $U^i$ were i.i.d. $N(0, 1)$ independent of $(X_1^i, \ldots, X_4^i)$. In terms of the representation (1.1), the centred coefficient functions, denoted by $f_{jk}(\cdot; \alpha)$, in the conditional $\alpha$-quantile function $m$ are thus given by

$$f_{13}(u; \alpha) = c_{13} + 4(u - 0.5)^2 + \exp(-u)\Phi^{-1}(\alpha),$$

$$f_{23}(u; \alpha) = c_{23} + \exp(u - 1),$$

$$f_{24}(u; \alpha) = c_{24} + \cos(2\pi u),$$

where the constants $c_{jk}$ are chosen so that $\int f_{jk}(u; \alpha) du = 0$, and $\Phi^{-1}(\alpha)$ is the $\alpha$-quantile of the standard normal distribution. The sample sizes were $n = 400$ and $n = 1000$.

We have used the R function `rq( )` in the library quantreg to optimize the objective functions at (2.1) and (2.2). For SBF, we discretized the integrals on a fine grid in $[0, 1]^2$. We used the Epanechnikov kernel given by $K(u) = (3/4)(1 - u^2)I_{[-1,1]}(u)$. For the bandwidths, we took a number of combinations $(h_1, h_2)$ with $h_1 \in [0.10, 0.40]$ and $h_2 \in [0.05, 0.20]$. We chose the
initial estimates in the iterations (2.1) and (2.2) to be zero. We found that the algorithms converged with this initial choice in all cases.

Tables 1 and 2 provide the Monte Carlo estimates of the mean integrated squared errors (MISE), the integrated squared biases (ISB) and the integrated variance (IV) of the BF and SBF estimators of the individual coefficient functions. Table 1 is for the case where the covariates $X_j$ are uncorrelated, while Table 2 is for the correlated case. In the tables, we only report the results for the bandwidth choice $(h_1, h_2) = (0.3, 0.15)$ in the case $n = 400$ and $(h_1, h_2) = (0.2, 0.1)$ in the case $n = 1000$, which give an average performance. The results were based on 500 pseudo-samples.

A close investigation of the tables reveals that the SBF estimator has slightly better MISE performance than the BF estimator in all cases. Note that the true coefficient functions $f_{23}$ and $f_{24}$ remain the same, as the level of quantile ($\alpha$) changes. Because of this, the values of the ISB of the estimators $\hat{f}_{23}^{BF}$, $\hat{f}_{23}^{SBF}$, $\hat{f}_{24}^{SBF}$ do not change much across different values of $\alpha$. We also find that both the BF and SBF methods show fairly good performance in the case of correlated covariates. Comparing the values of MISE, ISB and IV in Table 2 with those in Table 1, we observe that the values of MISE, ISB and IV in the correlated case are marginally larger than the corresponding values in the uncorrelated case.

ACKNOWLEDGEMENTS

This research was supported by the National Research Foundation of Korea (NRF) Grant (NRF-2010-0021396; Y. K. Lee), by the Collaborative Research Center SFB 884 “Political Economy of Reforms”, financed by the German Science Foundation (DFG; Enno Mammen) and by the NRF Grant funded by the Korea government (MEST) (No. 2010-0017437; B. U. Park).

REFERENCES

Buja, A., T. Hastie and R. Tibshirani (1989). Linear smoothers and additive models (with discussion). *Annals of Statistics* 17, 453–555.

Cai, Z., M. Das, H. Xiong and X. Wu (2006). Functional coefficient instrumental variables models. *Journal of Econometrics* 133, 207–41.

Furno, M. (2004). ARCH tests and quantile regressions. *Journal of Statistical Computation and Simulation* 74, 277–92.

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

Honda, T. (2004). Quantile regression in varying coefficient models. *Journal of Statistical Planning and Inference* 121, 113–25.

Hong, Y. and T-H. Lee (2003). Inference and forecast of exchange rates via generalized spectrum and nonlinear time series models. *Review of Economics and Statistics* 69, 1048–62.

Horowitz, J. and S. Lee (2005). Non-parametric estimation of an additive quantile regression model. *Journal of the American Statistical Association* 100, 1238–49.

Kim, M-O. (2007). Quantile regression with varying coefficients. *Annals of Statistics* 35, 92–108.

Lee, Y. K., E. Mammen and B. U. Park (2010). Backfitting and smooth backfitting for additive quantile models. *Annals of Statistics* 38, 2857–83.

Lee, Y. K., E. Mammen and B. U. Park (2012a). Projection-type estimation for varying coefficient regression model. *Bernoulli* 18, 177–205.
Lee, Y. K., E. Mammen and B. U. Park (2012b). Flexible generalized varying coefficient regression models. *Annals of Statistics* 40, 1906–33.

Li, Q., C. Huang and F. Fu (2002). Semiparametric smooth coefficient model. *Journal of Business and Economic Statistics* 20, 412–22.

Linton, O. and J. P. Nielsen (1995). A kernel method of structured non-parametric regression based on marginal integration. *Biometrika* 83, 529–40.

Mammen, E., O. Linton and J. P. Nielsen (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics* 27, 1443–90.

Mammen, E. and B. U. Park (2005). Bandwidth selection for smooth backfitting in additive models. *Annals of Statistics* 33, 1260–94.

Mammen, E. and B. U. Park (2006). A simple smooth backfitting method for additive models. *Annals of Statistics* 34, 2252–71.

Opsomer, J. D. and D. Ruppert (1997). Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics* 25, 186–211.

Park, B. U., E. Mammen, Y. K. Lee and E. R. Lee (2013). Varying coefficient regression models: a review and new developments. Forthcoming in *International Statistical Review* (doi:10.1111/insr.12029).

Roca-Pardinas, J. and S. Sperlich (2010). Feasible estimation in generalized structured models. *Statistics and Computing* 20, 367–79.

van de Geer, S., (2000). *Empirical Processes in M-Estimation*. Cambridge: Cambridge University Press.

Wang, H. J., Z. Zhu and J. Zhou (2009). Quantile regression in partially linear varying coefficient models. *Annals of Statistics* 37, 3841–66.

Xue, L. and L. Yang (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica* 16, 1423–46.

Yang, L., B. U. Park, L. Xue and W. Härdle (2006). Estimation and testing for varying coefficients in additive models with marginal integration. *Journal of the American Statistical Association* 101, 1212–27.

Yu, K., B. U. Park and E. Mammen (2008). Smooth backfitting in generalized additive models. *Annals of Statistics* 36, 228–60.

**APPENDIX: PROOFS OF RESULTS**

**Proof of Theorem 3.1:** The proof makes essential use of the theory developed in Lee et al. (2010, 2012b). Lee et al. (2012b) considered a generalized varying coefficient model for mean regression. Lee et al. (2010) developed asymptotic theory for the ordinary BF and the SBF in non-parametric additive quantile regression. The basic idea in Lee et al. (2010) was to apply Bahadur expansions to show that the backfitting quantile estimators behave, in the first-order, like weighted backfitting estimators in a hypothetical additive mean regression model. Here, we outline how this theory can be adapted to our setting. We give some details for the ordinary BF. The proof for the SBF follows along similar lines.

We consider the following hypothetical model

\[ Z^i = m(X^i) + \eta^i = \tilde{X}_1^i f_1(X_{r+1}^i) + \cdots + \tilde{X}_p^i f_p(X_{r+p}^i) + \eta^i \]

with

\[ \eta^i = - \frac{I(e^i \leq 0) - \alpha}{\rho_{e^i X}(0|X^i)}. \]
For this model, we define a linear backfitting estimator $\hat{f}_j^{\text{BF}}(u)$ that is based on the following iterations,

$$
\hat{f}_j^{\text{BF},[k+1]}(u) = \arg\min_{\theta \in \Theta_j} \sum_{i=1}^{n} (Z_i - \hat{X}_j^\top \theta - \sum_{\ell=1}^{j-1} \hat{X}_\ell^\top \hat{f}_{[\ell]}^{\text{BF},[k+1]}(X_{r+\ell}))^2
$$

$$
- \sum_{\ell=j+1}^{p} \hat{X}_\ell^\top \hat{f}_{\ell}^{\text{BF},[k]}(X_{r+\ell}))^2 p_j(X(0)|X') K_{h_j}(u, X_{r+\ell}),
$$

(A.1)

with starting values $\hat{f}_j^{\text{BF},[0]} = \hat{f}_{[0]}^{\text{BF}}$. Similarly, we can define a linear SBF estimator $\hat{f}_j^{\text{SBF}}(u)$ based on the iterations:

$$
\hat{f}_j^{\text{SBF},[k+1]}(u) = \arg\min_{\theta \in \Theta_j} \int (Z_i - \hat{X}_j^\top \theta - \sum_{\ell=1}^{j-1} \hat{X}_\ell^\top \hat{f}_{[\ell]}^{\text{SBF},[k]}(z_\ell))^2 p_j(X(0)|X') \prod_{\ell=1, \neq j} K_{h_j}(z_\ell, X_{r+\ell}) \, dz_\ell
$$

$$
\times K_{h_j}(u, X_{r+\ell}).
$$

(A.2)

The latter estimator differs from the estimator considered in Lee et al. (2012b) by the additional weight $p_j(X(0)|X')$ in (A.2). We can show that the theory developed by Lee et al. (2012b) can be extended to the case of an additional weight. In particular, under the assumptions of Theorem 3.1, it holds that for fixed $n$ and for $k \to \infty$ the estimator $\hat{f}_j^{\text{SBF},[k]}$ converges with a geometric speed to a limit $\hat{f}_j^{\text{SBF}}$, which is the minimizer of

$$
\sum_{i=1}^{n} (Z_i - \sum_{j=1}^{p} \hat{X}_j^\top \hat{f}_j^{\text{SBF}}(z_j))^2 p_j(X(0)|X') \prod_{j=1}^{p} K_{h_j}(z_j, X_{r+j}) \, dz_j.
$$

(A.3)

With probability tending to one, the convergence of the algorithm is uniform over $n$; compare Theorem 4 in Lee et al. (2012b). As in Theorem 3 of Lee et al. (2012b), we can show that for all $z$ in the interior of the support of $p_X$,

$$
n^{2/5} (\hat{f}_j^{\text{SBF},[k]}(z_j) - f_j(z_j)) \text{ are jointly asymptotically normal with mean} \left(\beta_1(z_j)^\top, \ldots, \beta_p(z_j)^\top\right)\text{ and variance } \text{diag}(\Sigma_j(z_j)).
$$

The same asymptotic normal limit applies for $n^{2/5} (\hat{f}_j^{\text{SBF},[k]}(z_j) - f_j(z_j))$ with $k^* = C_{\text{iter,SBF}} \log n$ if $C_{\text{iter,SBF}}$ is chosen large enough or if $|\hat{f}_j^{[0]}(z_j) - f_j(z_j)|$ is small enough (Assumption 3.5). This follows from the uniform geometric convergence of $\hat{f}_j^{\text{SBF},[k]}$ for $k \to \infty$. According to the assumptions of Theorem 3.1, these conditions can be satisfied.

For the statement of the theorem on the ordinary BF estimator, we show that

$$
\hat{f}_j^{\text{BF},[k^*]}(u) = \hat{f}_j^{\text{BF}}(u) + o_p(n^{-2/5}),
$$

(A.4)

$$
n^{2/5} (\hat{f}_j^{\text{BF},[k^*]}(u) - f_j(u)) \to N\left(\left(\beta_1^{\text{BF}}(z_j)^\top, \ldots, \beta_p^{\text{BF}}(z_j)^\top\right)^\top, \text{diag}(\Sigma_j(z_j))\right)
$$

(A.5)

in distribution. For the asymptotic normality result on the SBF estimator, we have to show that $\hat{f}_j^{\text{SBF},[k^*]}(u) = \hat{f}_j^{\text{SBF}}(u) + o_p(n^{-2/5})$ and that

$$
n^{2/5} (\hat{f}_j^{\text{SBF},[k^*]}(u) - f_j(u)) \to N\left(\left(\beta_1^{\text{SBF}}(z_j)^\top, \ldots, \beta_p^{\text{SBF}}(z_j)^\top\right)^\top, \text{diag}(\Sigma_j(z_j))\right)
$$

in distribution. The latter two claims follow with similar arguments as (A.4) and (A.5), and for this reason their proofs are omitted. We only outline why the bias terms for the ordinary BF and for the SBF differ.

For the proof of (A.4), we can proceed similarly as in the proof of Proposition 2.1 in Lee et al. (2010). Using a similar Bahadur expansion, we can show that with an appropriate choice of $\delta > 0$ we obtain the
following implication. Assume for a $k \leq k^*$ that
\begin{equation}
\sup_{C_{sh_j} \leq u \leq 1 - C_{sh_j}} \left| \hat{\hat{f}}^+,_{BF,[k]}(u) - \hat{f}^+,_{BF,[k]}(u) \right| \leq n^{-5(1/2) - \delta} R_{k,n}, \tag{A.6}
\end{equation}
\begin{equation}
\sup_{0 \leq u \leq 1} \left| \hat{\hat{f}}^+,_{BF,[k]}(u) - \hat{f}^+,_{BF,[k]}(u) \right| \leq n^{-5(1/2) - \delta} R_{k,n}, \tag{A.7}
\end{equation}
\begin{equation}
\sup_{C_{sh_j} \leq u \leq 1 - C_{sh_j}} \left| \hat{\hat{f}}^+,_{BF,[k]}(u) - g_{j,k}(u) \right| \leq n^{-5(1/2) - \delta} R_{k,n}, \tag{A.8}
\end{equation}
\begin{equation}
\sup_{0 \leq u \leq 1} \left| \hat{\hat{f}}^+,_{BF,[k]}(u) - g_{j,k}(u) \right| \leq n^{-5(1/2) - \delta} R_{k,n} \tag{A.9}
\end{equation}
for some random functions $g_{1,k}, \ldots, g_{p,k}$ with derivatives that fulfil the Lipschitz condition
\[|g'_{j,k}(u) - g'_{j,k}(v)| \leq C |u - v|^{\rho} n^\delta\]
for a constant $C$ that does not depend on $n$, and for $j = 1, \ldots, p$ and $u, v \in [0, 1]$. Here, $R_{k,n}$ is a random variable with $R_{k,n} = O_p(1)$. Using similar arguments as in the proof of Proposition 2.1 in Lee et al. (2010), we can show that (A.6)–(A.9) imply that
\begin{equation}
\sup_{C_{sh_j} \leq u \leq 1 - C_{sh_j}} \left| \hat{\hat{f}}^+,_{BF,[k+1]}(u) - \hat{f}^+,_{BF,[k+1]}(u) \right| \leq n^{-5(1/2) - \delta} R_{k+1,n}, \tag{A.6+}
\end{equation}
\begin{equation}
\sup_{0 \leq u \leq 1} \left| \hat{\hat{f}}^+,_{BF,[k+1]}(u) - \hat{f}^+,_{BF,[k+1]}(u) \right| \leq n^{-5(1/2) - \delta} R_{k+1,n}, \tag{A.7+}
\end{equation}
\begin{equation}
\sup_{C_{sh_j} \leq u \leq 1 - C_{sh_j}} \left| \hat{\hat{f}}^+,_{BF,[k+1]}(u) - g_{j,k+1}(u) \right| \leq n^{-5(1/2) - \delta} R_{k+1,n}, \tag{A.8+}
\end{equation}
\begin{equation}
\sup_{0 \leq u \leq 1} \left| \hat{\hat{f}}^+,_{BF,[k+1]}(u) - g_{j,k+1}(u) \right| \leq n^{-5(1/2) - \delta} R_{k+1,n} \tag{A.9+}
\end{equation}
for some random functions $g_{1,k+1}, \ldots, g_{p,k+1}$ with derivatives that fulfil the Lipschitz condition
\[|g'_{j,k+1}(u) - g'_{j,k+1}(v)| \leq C |u - v|^{\rho} n^\delta\]
for a $C$ as above, and for $j = 1, \ldots, p$ and $u, v \in [0, 1]$. Here, $R_{k+1,n}$ is again a random variable with $R_{k+1,n} = O_p(1)$.

Using this induction argument and Assumptions 3.5 and 3.6, we find that (A.6)–(A.9) hold for all $k \leq k^*$. Moreover, arguing similarly as in Lee et al. (2010), we can show that for some constant $\rho > 1$ it holds that
\[\sup_{k \geq 0} \rho^{-k} R_{k,n} = O_p(1).\]
This implies (A.4) by assuming that $\delta$ is large enough and that $C_{iter,SBF}$ is small enough.
To prove (A.5), we note that the solution of the minimization at (A.1) has the following explicit form:

\[
\hat{f}_{j,BF,[k+1]}(u) = \left( \sum_{i=1}^{n} \tilde{X}_{j}^{l}^{T} p_{c,i}(X|\hat{X}^{l}) K_{h_{j}}(u, X_{r+j}) \right)^{-1} \times \left( \sum_{i=1}^{n} \tilde{X}_{j}^{l} Z_{i} p_{c,i}(X|\hat{X}^{l}) K_{h_{j}}(u, X_{r+j}) \right)
\]

\[- \sum_{i=1}^{n} \sum_{\ell=1}^{p} \hat{f}_{i,BF,[k+1]}(X_{r+j}^{i}) p_{c,i}(X|\hat{X}^{l}) K_{h_{j}}(u, X_{r+j}^{i}) \]

\[- \sum_{i=1}^{n} \sum_{j=1}^{p} \hat{X}_{j}^{l}^{-1} \hat{f}_{i,BF,[k+1]}(X_{r+j}^{i}) p_{c,i}(X|\hat{X}^{l}) K_{h_{j}}(u, X_{r+j}^{i}). \]

We rewrite this as

\[
\hat{f}_{j,BF,[k+1]}(u) = f_{j}(u) + \hat{f}_{j}^{A}(u) + \hat{f}_{j}^{B}(u)
\]

\[- (\hat{W}_{j}(u))^{-1} n^{-1} \sum_{i=1}^{n} \hat{X}_{j}^{l} \hat{X}_{j}^{l}^{T} \hat{f}_{i,BF,[k+1]}(X_{r+j}^{i}) - f_{i}(X_{r+j}^{i}) \]

\[\times p_{c,i}(X|\hat{X}^{l}) K_{h_{j}}(u, X_{r+j}^{i}) \]

where \(k_{\ell,j} = k + I(j < \ell)\) and

\[\hat{W}_{j}(u) = n^{-1} \sum_{i=1}^{n} \hat{X}_{j}^{l} \hat{X}_{j}^{l}^{T} p_{c,i}(X|\hat{X}^{l}) K_{h_{j}}(u, X_{r+j}^{i}),\]

\[\hat{f}_{j}^{A}(u) = n^{-1} \hat{W}_{j}^{-1}(u) \sum_{i=1}^{n} \hat{X}_{j}^{l} n^{j} p_{c,i}(X|\hat{X}^{l}) K_{h_{j}}(u, X_{r+j}^{i}),\]

\[\hat{f}_{j}^{B}(u) = n^{-1} \hat{W}_{j}^{-1}(u) \sum_{i=1}^{n} \hat{X}_{j}^{l} \hat{X}_{j}^{l}^{T} \left( f_{j}(X_{r+j}^{i}) - f_{j}(u) \right) p_{c,i}(X|\hat{X}^{l}) K_{h_{j}}(u, X_{r+j}^{i}).\]

Then, from (A.6)–(A.9), for \(\delta^{*} > 0\) small enough, we obtain

\[\left( \hat{W}_{j}(u) \right)^{-1} n^{-1} \sum_{i=1}^{n} \hat{X}_{j}^{l} \hat{X}_{j}^{l}^{T} \hat{f}_{i,BF,[k+1]}(X_{r+j}^{i}) p_{c,i}(X|\hat{X}^{l}) K_{h_{j}}(u, X_{r+j}^{i}) \]

\[= \int (\hat{W}_{j}(u))^{-1} \hat{W}_{j}(u, z)(\hat{f}_{i,BF,[k+1]}(z) - f_{j}(z)) dz + o_{p}(n^{-(2/5) - \delta^{*}}). \quad (A.11)\]

uniformly for \(u \in [0, 1]\) and for \(1 \leq k \leq k^{*}, 1 \leq j, \ell \leq p, \ell \neq j\). This follows from Lemma A.1 and the fact that

\[\sup_{u \in [0, 1]} |\hat{W}_{j}(u) - W_{j}(u)| = O_{p}(n^{-2/5} \sqrt{\log n}). \quad (A.12)\]

For (A.12), we have used the fact that \(K_{h_{j}}\) is a kernel with boundary correction so that it fulfills \(\int K_{h_{j}}(u, x) dx = 1\) and \(\int K_{h_{j}}(u, x)(u - x) dx = 0\).

Define \(\hat{\beta}_{j}^{B}(u) = (1/2)c_{j}^{2} \hat{f}_{j}^{B}(u) \int t^{2} K(t) dt\) and recall the definition of \(\hat{\beta}_{j}^{B}(u)\) given below Assumption 3.7. With starting values \(\hat{f}_{j}^{A,0}(u) \equiv 0\) and \(\hat{f}_{j}^{B,0}(u) = \hat{f}_{j}^{A,0}(u) - f_{j}(u) = \hat{f}_{j}^{B,0}(u) - f_{j}(u)\), consider the
following iterations to obtain $\hat{f}^{A,[k]}_j$ and $\hat{f}^{B,[k]}_j$ for $k \geq 1$:

$$
\hat{f}^{A,[k+1]}_j(u) = \hat{f}^{A}_j(u) - \sum_{\ell=1,\neq j}^{p} \int (W_{\beta}(u))^{-1} W_{j\ell}(u, z) \hat{f}^{A,[k]}_{\ell}(z) \, dz,
$$

$$
\hat{f}^{B,[k+1]}_j(u) = n^{-2/5} \beta^{*}_j(u) + n^{-2/5} \beta^{**}_j(u) - \sum_{\ell=1,\neq j}^{p} \int (W_{\beta}(u))^{-1} W_{j\ell}(u, z) \hat{f}^{B,[k]}_{\ell}(z) \, dz.
$$

The update $\hat{f}^{A,[k]}_j$ turns out to be the leading stochastic term of $\hat{f}^{+,BF,[k]}_j$, and $\hat{f}^{B,[k]}_j$ makes the main contribution of the bias of $\hat{f}^{+,BF,[k]}_j$. See also the above definition of $\hat{f}^{A}_j$. From (A.11) and Lemma A.2, we obtain

$$
\sup_{C_{h,j} \leq u \leq 1-C_{h,j}} |\hat{f}^{+,BF,[k]}_j(u) - f_j(u) - \hat{f}^{A,[k]}_j(u) - \hat{f}^{B,[k]}_j(u)| = o_p(n^{-2/5}),
$$

$$
\sup_{0 \leq u \leq 1} |\hat{f}^{+,BF,[k]}_j(u) - f_j(u) - \hat{f}^{A,[k]}_j(u) - \hat{f}^{B,[k]}_j(u)| = o_p(n^{-2/5}).
$$

Arguing as in other works on BF or SBF, it can be shown that

$$
\sup_{0 \leq u \leq 1} |\hat{f}^{A,[k]}_j(u) - \hat{f}^{A}_j(u)| = o_p(n^{-2/5}).
$$

The main argument here is that terms of the type $\int (W_{\beta}(u))^{-1} W_{j\ell}(u, z) \hat{f}^{A,[k]}_{\ell}(z) \, dz$ are of order $n^{-1/2}$ because global averages of local averages behave like global averages.

Put $\beta^{(3)}_j(u)$ to be the normalized versions of $n^{1/5} \hat{f}^{B,[k]}_j(u)$ so that they satisfy the constraints (1.2). We now argue that for $k \to \infty$, $\beta^{(3)}_j(u)$ converges to $\beta^{BF}_j(u)$. This can be shown as in the proof of Theorem 4 in Lee et al. (2012b). Moreover, as argued there in a only slightly different setting, it holds that, with probability tending to one,

$$
\sum_{j=1}^{p} \sum_{\ell=1}^{p} \int (\beta^{(3)}_j(z) - \beta^{BF}_j(z)) \top W_{j\ell}(z_j, z_\ell) (\beta^{(3)}_\ell(z_\ell) - \beta^{BF}_\ell(z_\ell)) \, dz_j \, dz_\ell \leq c \, n^{2\delta_1} k^k
$$

for some constant $c > 0$, $0 < \kappa < 1$ and $\delta_1 > 0$, which is slightly larger than $\delta_1$, where $\delta_1$ is the constant in Assumption 3.5. Here, we interpret $W_{\beta}(z_j, z_\ell)$ to be $W_{\beta}(z_j)$. Because of

$$
\beta^{(k+1)}_j(u) - \beta^{BF}_j(u) = \sum_{\ell=1,\neq j}^{p} \int (W_{\beta}(u))^{-1} W_{j\ell}(u, z) (\beta^{(k)}_\ell(z_\ell) - \beta^{BF}_\ell(z_\ell)) \, dz_j \, dz_\ell,
$$

this implies that $\sup_{0 \leq u \leq 1} |\beta^{(k)}_j(u) - \beta^{BF}_j(u)| \leq c^* n^{2\delta_1} k^k$ with probability tending to one, for some constant $c^*$. In particular, we obtain $\sup_{0 \leq u \leq 1} |\beta^{(k)}_j(u) - \beta^{BF}_j(u)| = o_p(1)$ if $C_{iter,BF}$ is large enough. Thus, we obtain

$$
\hat{f}^{+,BF,[k]}_j(u) = f_j(u) + n^{-2/5} \beta^{BF}_j(u) + \hat{f}^{A}_j(u) + o_p(n^{-2/5}).
$$

From this expansion, we obtain claim (A.5) by standard smoothing theory.

We now briefly outline the bias expression in the case of the SBF. Instead of (A.10), here we have

$$
\hat{f}^{+,SBF,[k+1]}_j(u) = f_j(u) + \hat{f}^{A}_j(u) + \hat{f}^{B}_j(u)
$$

$$
- (W_{\beta}(u))^{-1} n^{-3} \sum_{i=1}^{n} \sum_{\ell=1,\neq j}^{p} \int X_i^\top X_\ell (\hat{f}^{+,SBF,[k]}_{\ell}(z) - f_{\ell}(X_{\ell,t}^\top)) \, dz 
$$

$$
\times p_X(0|X') K_{\beta_j}(u, X_{\ell,t}) K_{\beta_\ell}(z, X_{\ell,t}) \, dz.
$$

(A.13)
Recall the definitions of \( \beta_{ij}^*(z_j) \) and \( \beta_{jj}^*(z_j) \) given below Assumption 3.7 and immediately after (A.17), respectively. Neglecting the boundary regions and the norming conditions (1.2), we can write (A.13) as

\[
\hat{\beta}^{SBF, \ell+1}\beta_j(u) - f_j(u) - \hat{\beta}^\ell_j(u) - n^{-2/5} \beta_{jj}^*(u)
\]

\[
\approx n^{-2/5} \beta_{jj}^*(u) - (W^\ell_j(u))^{-1} n^{-1} \sum_{i=1}^{n} \sum_{\ell', \ell''} \int \tilde{X}_i \tilde{X}_i^T (\hat{\beta}^{SBF, \ell_{i+1}} - f_i(z)) - f_i(z)
\]

\[
- (f_i(X_{r+1}^i) - f_i(z)) p_c(X(0)|X') K_{j}(u, X_{r+1}) K_{j}(z, X_{r+1}^i) d\zeta.
\]  

(A.14)

Approximating further the integral terms involving \( (f_i(X_{r+1}^i) - f_i(z)) \) on the right-hand side, we obtain

\[
n^{-1} \sum_{i=1}^{n} \int \tilde{X}_i \tilde{X}_i^T (f_i(X_{r+1}^i) - f_i(z)) p_c(X(0)|X') K_{j}(u, X_{r+1}) K_{j}(z, X_{r+1}^i) d\zeta
\]

\[
\approx \int W_j(u, z) n^{-2/5} \beta_j^*(z) d\zeta + n^{-2/5} \beta_{jj}^*(u).
\]

Taking this approximation into (A.14) and using the fact that \( \int (W^\ell_j(u))^{-1} W_j(u, z) \hat{\beta}^\ell_j(z) d\zeta \) are of order \( n^{-1/2} \), we obtain

\[
\hat{\beta}^{SBF, \ell+1}\beta_j(u) - f_j(u) - \hat{\beta}^\ell_j(u) - n^{-2/5} \beta_{jj}^*(u)
\]

\[
\approx \sum_{\ell=1}^{p} n^{-2/5} \beta_j^*(u) - \sum_{\ell=1}^{p} (W^\ell_j(u))^{-1} \int W_j(u, z)
\]

\[
\times (\hat{\beta}^{SBF, \ell_{i+1}} - f_i(z) - \hat{\beta}^\ell_j(z) - n^{-2/5} \beta_{jj}^*(z)) d\zeta.
\]

This entails

\[
\hat{\beta}^{SBF, \ell+1}(u) \approx f_j(u) + \hat{\beta}^\ell_j(u) + n^{-2/5} \beta_j^*(u) + n^{-2/5} \beta_{jj}^*(u).
\]

Lemma A.1. \textit{Let} \( \mathcal{M}_n \equiv M_n(C_1, C_2, \delta^+) \) \textit{denote a class of functions} \( \psi : [0, 1] \rightarrow \mathbb{R} \) \textit{such that} \( |\psi'(u) - \psi'(v)| \leq C_1 |u - v|^{\alpha} \) \textit{for all} \( u, v \in [0, 1] \), \( |\psi(u)| \leq C_2 n^{-1/5} \) \textit{for all} \( u \in [C_2 n^{-1/5}, 1 - C_2 n^{-1/5}] \) \textit{and} \( |\psi(u)| \leq n^{1/5} \gamma_n \) \textit{for all} \( u \in [0, 1] \), \textit{where} \( \gamma_n = n^{-2(3/5)\delta^+} \). \textit{Under the conditions of Theorem 3.1, there exist constants} \( \delta^+ > 0 \) \textit{and} \( \delta^+ > 0 \) \textit{such that for all} \( C_1, C_2 > 0 \)

\[
\sup_{g \in \mathcal{M}_n} \sup_{u \in [0, 1]} |n^{-1} \int \tilde{X}_i \tilde{X}_i^T g(X_{r+1}^i) p_c(X(0)|X') K_{j}(u, X_{r+1}) d\zeta| - \int W_j(u, z) g(z) d\zeta| \leq o_p(n^{-(2/5)\delta^+}).
\]

\textbf{Proof:} \textit{We first note that, uniformly for} \( u, z \in [0, 1] \),

\[
n^{-1} \sum_{i=1}^{n} W_j(X_{r+1}^i, z) K_{j}(u, X_{r+1}^i) p_j(X_{r+1})^{-1} - W_j(u, z) = O_p(n^{-2/5} \sqrt{\log n}).
\]

This follows from the standard technique of kernel smoothing and from the condition on the kernel function that \( \int K_{j}(u, x) dx = 1 \) and \( \int K_{j}(u, x)(u - x) dx = 0 \). Thus, it suffices to prove

\[
\sup_{g \in \mathcal{M}_n} \sup_{u \in [0, 1]} |n^{-1} \int D^\ell_j(u, g) | = o_p(n^{-(2/5)\delta^+}),
\]

(A.15)
where

\[
D_{ij}^j(u, g) = K_h_j(u, X_{r+j}^i) \left( \bar{X}_j^i \bar{X}_j^i X_{r+j}^i \right) \tilde{p}_j(x(0)X) g(X_{r+j}^i) \\
- E\left[ \bar{X}_j^i \bar{X}_j^i X_{r+j}^i \tilde{p}_j(x(0)X) g(X_{r+j}^i) | X_{r+j}^i \right].
\]

The property (A.15) can be proved by an application of the empirical process theory given by van de Geer (2000). Let \( \mathcal{M}_n (2^{-j}) \) be a grid of points in \( \mathcal{M}_n \) such that for every \( g \in \mathcal{M}_n \) there exists \( g_j \in \mathcal{M}_n (2^{-j}) \) with \( \| g - g_j \|_\infty \leq 2^{-j} \gamma_n \), where \( \| \cdot \|_\infty \) denotes the sup-norm. For each \( g \in \mathcal{M}_n \), choose such a point \( g_j \equiv g_j(g) \in \mathcal{M}_n (2^{-j}) \) for \( j \geq 1 \), and take \( g_0 \equiv 0 \). Define

\[
J_n = \min\{ j \geq 1 : 2^{-j} < n^{-\delta^+} \},
\]

where \( \delta^+ \) and \( \delta^+_1 \) are positive constants in a region to be specified later. Then, we obtain

\[
\sup_{g \in \mathcal{M}_n} \sup_{u \in [0,1]} |n^{-1} \sum_{i=1}^{n} (D_{ij}^j(u, g) - D_{ij}^j(u, g_{j+1}))| = o_p(n^{-(2/5) - \delta^+_1}) \tag{A.16}
\]

because \( n^{-1} \sum_{i=1}^{n} K_h_j(u, X_{r+j}^i) = O_p(1) \) uniformly for \( u \in [0,1] \). Let \( N_j \) denote the \( 2^{-j} \gamma_n \) entropy of \( \mathcal{M}_n \) for the sup-norm. It is bounded by a constant multiple of \( n^{k/(1+\rho)}2^{j/(1+\rho)} \). Take \( \eta_j > 0 \) such that \( \sum_{j=1}^{J_n} \eta_j \leq 1 \). Then, writing \( \mathcal{X}_{r+j} = (X_{r+j}^1, \ldots, X_{r+j}^n) \), we find that, for any \( \kappa > 0 \)

\[
I \equiv P\left( \sup_{g \in \mathcal{M}_n} \sup_{u \in [0,1]} |n^{-1} \sum_{i=1}^{n} (D_{ij}^j(u, g) - D_{ij}^j(u, g_{j+1}))| > \kappa n^{-(2/5) - \delta^+_1} | \mathcal{X}_{r+j} \right)
\]

\[
\leq \sum_{k=1}^{J_n} \exp(N_k + N_{k-1}) \\
\times \max_{u \in [0,1]} P\left( \sup_{g \in \mathcal{M}_n} |n^{-1} \sum_{i=1}^{n} (D_{ij}^j(u, g_k) - D_{ij}^j(u, g_{k-1}))| > \kappa \eta_k n^{-(2/5) - \delta^+_1} | \mathcal{X}_{r+j} \right), \tag{A.17}
\]

where \( \max_n \) is taken over all \( g_k \in \mathcal{M}_n (2^{-k}) \) and \( g_{k-1} \in \mathcal{M}_n (2^{-k+1}) \) with \( \| g_k - g_{k-1} \|_\infty \leq \gamma_n \).

Now, we take a grid \( (u_k : 1 \leq k \leq n^G/2) \) on \([0,1]\) such that for every \( u \in [0,1] \) there exists \( u_k \) such that \( |u - u_k| \leq n^{-G} \). Take \( G \) sufficiently large so that

\[
|D_{ij}^j(u, g) - D_{ij}^j(v, g)| \leq \kappa \eta_k n^{-(2/5) - \delta^+_1}
\]

for all \( 1 \leq i \leq n, 1 \leq k \leq J_n, g \in \mathcal{M}_n \) and \( u, v \in [0,1] \) with \( |u - v| \leq n^{-G} \). A sufficiently large \( G \) exists because of the Lipschitz condition on the kernel function (see Assumption 3.2). Then, we obtain from (A.17)

\[
I \leq G \sum_{k=1}^{J_n} \exp(N_k + N_{k-1} + \log n) \\
\times \max_{u_k} \max_{g_k} P\left( |n^{-1} \sum_{i=1}^{n} (D_{ij}^j(u_k, g_k) - D_{ij}^j(u_k, g_{k-1}))| > \kappa \eta_k n^{-(2/5) - \delta^+_1} | \mathcal{X}_{r+j} \right). \tag{A.18}
\]
Applying the Hoeffding inequality to the probability at (A.18) conditioning on \( X_{r+j} \), we find that, for some constants \( C_3, C_4 > 0 \),

\[
1 \leq 2G \sum_{k=1}^{J_n} \exp \left( 2 \cdot 2^{k/(1+\rho)} n^{\xi/(1+\rho)} + \log n - 2C_3k^2 2^k \eta_k^2 n^{3/(5)-2\delta-2\delta_1^2} \right) \\
\leq 2G \sum_{k=1}^{J_n} \exp \left( -C_3k^2 2^k \eta_k^2 n^{3/(5)-2\delta-2\delta_1^2} \right) \\
\leq 2G \sum_{k=1}^{J_n} \exp \left( -kC_4k^2 n^{3/(5)-2\delta-2\delta_1^2} \right) \\
\leq 4G \exp \left( -C_4k^2 n^{3/(5)-2\delta-2\delta_1^2} \right) .
\]

In the second and third inequalities, we have used specifically

\[
\eta_k = \max \left\{ -k^{1/2} - k^{1/2} - k^{1/2} - k^{1/2} - \frac{\delta^+ + \delta_1^+ + \delta_1^+ + \delta_1^+ + \delta_1^+ + (3/10)}{n} \right\} .
\]

Thus, taking any \( \delta^+ \) and \( \delta_1^+ \) such that \( \xi/(1+\rho) + \delta^+ + \delta_1^+ < 3/10 \) suffices to prove \( E[I] \to 0 \), which together with (A.16) establishes (A.15). This completes the proof of Lemma A.1.

**Lemma A.2.** Under the conditions of Theorem 3.1, it holds that

\[
\sup_{C_3h_j \leq u \leq 1 - C_3h_j} \left| \mathbf{\hat{F}}_i^\theta(u) - n^{-2/5} \mathbf{\hat{B}}_i^\theta(u) - n^{-2/5} \mathbf{\hat{B}}_i^\theta(u) \right| = o_p(n^{-2/5}), \\
\sup_{0 \leq u \leq 1} \left| \mathbf{\hat{F}}_i^\theta(u) - n^{-2/5} \mathbf{\hat{B}}_i^\theta(u) - n^{-2/5} \mathbf{\hat{B}}_i^\theta(u) \right| = O_p(n^{-2/5}).
\]

**Proof:** Let \( \eta_j(v, u) = E[\mathbf{\hat{X}}_i \mathbf{\hat{X}}_i^\top (\mathbf{f}_j(v) - \mathbf{f}_j(u))p_{\epsilon}(0|\mathbf{X})|X_{r+j} = v] \). Then, it follows that

\[
n^{-1} \sum_{i=1}^{n} \left( \mathbf{\hat{X}}_i \mathbf{\hat{X}}_i^\top (\mathbf{f}_j(v) - \mathbf{f}_j(u))p_{\epsilon}(0|\mathbf{X}) - \eta_j(X_{r+j}, u) \right) K_{h_j}(u, X_{r+j}) \\
= O_p(n^{-3/5} \sqrt{\log n})
\]

uniformly for \( u \in [0, 1] \). Because the kernel \( K_{h_j} \) is boundary corrected as in Assumption 3.2, we obtain

\[
n^{-1} \sum_{i=1}^{n} \eta_j(X_{r+j}, u) K_{h_j}(u, X_{r+j}) = \eta_j(u, u) + \frac{1}{2} \left( \frac{\partial^2}{\partial v^2} \eta_j(v, u) p_j(v) \right)_{v=u} \\
\times \int_0^1 (v - u)^2 K_{h_j}(u, v) dv + o_p(n^{-2/5})
\]

uniformly for \( u \in [0, 1] \). Note that

\[
\int_0^1 (v - u)^2 K_{h_j}(u, v) dv = h_j^2 \int r^2 K(r) dr, \quad u \in [C_3h_j, 1 - C_3h_j]
\]

and it is \( O(h_j^2) \) for \( u \in [0, 1] \). The lemma now follows from (A.12) and the facts \( \eta_j(u, u) = 0, \Delta_j = 0 \) and

\[
\left( \frac{\partial^2}{\partial v^2} \eta_j(v, u) p_j(v) \right)_{v=u} = 2 \mathbf{f}_j(u) p_j(u) E[\mathbf{\hat{X}}_i \mathbf{\hat{X}}_i^\top (p_{\epsilon}(0|\mathbf{X}) - p_{\epsilon}(0|\mathbf{X}))|X_{r+j} = u] \\
+ \mathbf{f}_j(u) p_j(u) E[\mathbf{\hat{X}}_i \mathbf{\hat{X}}_i^\top p_{\epsilon}(0|\mathbf{X})|X_{r+j} = u].
\]