We propose a new estimation method for generalized varying coefficient models where the link function is specified up to some smoothness conditions. Consistency of the estimated varying coefficients and the link function estimate is established.

1 Estimation Method

Suppose we have independent and identically distributed observations \((Y_i, X_i, U_i)\) following the generalized varying coefficient model

\[
Y_i = g\{\sum_{k=1}^{p} \beta_k(U_i)X_{ik}\} + \epsilon_i
\]

with \(E(\epsilon_i \mid U_i, X_i) = 0\) \((i = 1, \ldots, n)\). Here \(g(\cdot)\) is called the link function and \(\beta_k(\cdot)\) s are the varying coefficients. The covariate \(X_i\) is p-dimensional and \(U_i\) is a univariate random variable and is called the effect modifier or the index variable.

For convenience, assume the coefficient functions \(\beta_k(\cdot)\) \((k = 1, \ldots, p)\) of model (1) are defined on \([0, 1]\) with each \(\beta_k(\cdot)\) having a continuous derivative. Our aim is to estimate these coefficients pointwise and therefore define for \(0 \leq u_0 \leq 1\) the local constant approximation of the linear predictor \(\eta(U_i) = \sum_{k=1}^{p} \beta_k(U_i)X_{ik}\) of model (1) as \(\eta(u_0) = \theta^T X_i\) where \(\theta = (a_1, \ldots, a_p)^T\) and \(X_i = (X_{i1}, \ldots, X_{ip})^T\). If the link function \(g\) were known, one could obtain local estimates of the coefficients by a straightforward weighted least squares minimization. Since \(g\) is unspecified in our model, a natural strategy is to estimate the link function nonparametrically. However, since the coefficient functions are unknown, the linear predictor is unknown and therefore we can not simply smooth the responses against the linear predictor to estimate the link function. Noting that \(\eta(u_0) = \theta^T X_i\) is a local approximation of the linear predictor, an estimate of the link function can be obtained by smoothing \(\{\theta^T X_i, Y_i\} \ (i = 1, \ldots, n)\) instead. However, if we simply use a standard one dimensional
smoother, we are ignoring the fact that $\eta_i(u_0)$ is a localized estimate of the true linear predictor for a given $U = u_0$ and as a result the link function estimate will be inconsistent as described in the following.

For example, let $\hat{g}_{nw}(t, \theta) = \sum_{i=1}^n w_i Y_i$ be the Nadaraya-Watson estimator of the link function at $t$ where $w_i = K_h(\theta^T X_i - t) / \sum_{j=1}^n K_h(\theta^T X_j - t)$ with $K_h(\cdot) = K(\cdot/h)$ is a symmetric kernel function. Now suppose we knew the true values of the coefficients of model (1) at $U = u_0$. Denote them as $\theta_0 = \{\beta_1(u_0), \ldots, \beta_p(u_0)\}^T$. Under suitable conditions on the smoothing parameter $h$ we know that $\hat{g}_{nw}(t, \theta_0) \to E(Y \mid \theta_0^T X = t)$ in probability as $n \to \infty$. However, $E(Y \mid \theta_0^T X = t) \neq g(t)$ and therefore, in order to consistently estimate the link function, we need to ensure that only the observations close to $u_0$ are being utilized in the smoothing process. To achieve this objective we introduce a Nadaraya-Watson type kernel smoother that consists of two kernels as our estimator of the link function.

For $t$ on $T$, the support of $\theta^T X$, let $\hat{g}_{u_0}(t, \theta) = A_{n,u_0}(t, \theta)/B_{n,u_0}(t, \theta)$ where

$$A_{n,u_0}(t, \theta) = (nh_1h_2)^{-1} \sum_{j=1}^n Y_j K_1\left(\frac{\theta^T X_j - t}{h_1}\right) K_2\left(\frac{U_j - u_0}{h_2}\right)$$

and $B_{n,u_0}(t, \theta)$ is $A_{n,u_0}(t, \theta)$ with $Y_j = 1$ $(j = 1, \ldots, n)$. Kernel $K_1(\cdot)$ with smoothing parameter $h_1$ localize the observations around the point of estimation $t$ as in the classical Nadaraya-Watson estimator. The kernel $K_2(\cdot)$ provides the additional restriction we need to achieve consistency by limiting the observations further into a neighborhood around $u_0$ determined by the smoothing parameter $h_2$. This estimator is similar in spirit to the estimator of Ichimura (1993). However, the weight function in his estimator does not depend on the sample size and is used as a means of handling heteroscedasticity. In contrast, the weights given by the extra kernel $K_2(\cdot)$ in our estimator depend on the sample size and serve the purpose of localizing the observations around the point $u_0$.

Given this estimator of the link function we propose to minimize the localized least squares criterion

$$M_n(\theta) = (nh_2)^{-1} \sum_{i=1}^n \{Y_i - \hat{g}_{u_0}(\theta^T X_i, \theta)\}^2 K_2\left(\frac{U_i - u_0}{h_2}\right)$$

(3)

to estimate the varying coefficients at $u_0$. The minimizer $\hat{\theta}$ of (3) estimates $\beta_k(u_0)$ $(k = 1, \ldots, p)$ and $\hat{g}_{u_0}(t, \hat{\theta})$ estimates the link function.

In what follows we will establish consistency results regarding this link function estimator and the coefficient function estimator.

### 2 Consistency of the Link Function Estimator

For a fixed $\theta \in \Theta$, let $T$ be the support of $\theta^T X$ for $X \in \mathcal{X}$. For $t \in T$, let $\hat{g}_{u_0}(t, \theta)$ be an estimator of the link function $g$, where $\hat{g}_{u_0}(t, \theta) = A_{n,u_0}(t, \theta)/B_{n,u_0}(t, \theta)$ where

$$A_{n,u_0}(t, \theta) = (nh_1h_2)^{-1} \sum_{j=1}^n Y_j K_1\left(\frac{\theta^T X_j - t}{h_1}\right) K_2\left(\frac{U_j - u_0}{h_2}\right)$$
and \( B_{n,u_0}(t, \theta) \) is \( A_{n,u_0}(t, \theta) \) with \( Y_j = 1 \) \((j = 1, \ldots, n)\).

To facilitate our arguments we impose the following technical conditions.

C1: The link function \( g \) and the coefficient functions \( \beta_k(\cdot) \) \((k = 1, \ldots, p)\) are three times continuously differentiable and \( g \) is non-constant on the support of \( \theta^T X \).

C2: The point \( \theta_0 = \{ \beta_1(u_0), \ldots, \beta_p(u_0) \}^T \) is an interior point of a compact set \( \Theta \).

C3: With an unspecified link, model (1) is not identifiable. Therefore, we constrain the coefficient functions such that \( \beta_1(u) > 0 \) and \( \sum_{k=1}^{p} \beta_k(u) = 1 \) for \( 0 \leq u \leq 1 \).

For each given \( U = u_0 \), with an unspecified link, model (1) is a single index model. This condition is the standard restriction in single index models (Ichimura, 1993; Lin & Kulasekera, 2007). It ensures that our objective function (3) has a well separated minimum (van der Vaart, 1998) in the neighborhood of the true parameter.

C4: There exist a positive definite matrix \( M_2(\theta_0) \) such that
\[
E \left\{ \frac{\partial^2 M_n(\theta)}{\partial \theta \partial \theta^T} \right\}_{\theta=\theta_0} \rightarrow M_2(\theta_0)
\]
as \( n \rightarrow \infty \) where \( M_2(\theta_0) \) is analogous to the information matrix in classical linear models.

This condition is similar to condition M7 of Chiou & Müller (1998) and condition 3 of Lemma 5.4 in Ichimura (1993).

C5: As \( n \rightarrow \infty \), \( h_1 \sim n^{-\delta_1} \), \( h_2 \sim n^{-\delta_2} \) with \( 0 < \delta_1 \leq 1/5 < \delta_2 < 1 \).

C6: The response variable is continuous with \( E (|Y|^m) < \infty \), for some \( m > 1 + (1 + 3\delta_1 + 2\delta_2)/(1 - 3\delta_1 - \delta_2) \). If we set the smoothing parameter \( h_1 \sim n^{-1/5} \), which is the optimal order for nonparametric regression functions estimators, we then require \( m > 6 \). We further assume the covariate \( X = (x_1, \ldots, x_p)^T \) satisfies \( \max_{1 \leq k \leq p} |x_k| \leq 1 \).

C7: The kernel functions \( K_1(\cdot) \), \( K_2(\cdot) \) are symmetric densities that are supported on \([-1, 1]\) and are continuously differentiable.

C8: Let \( A^{(k)}(X, \theta) \) be the \( k \)th partial derivative of \( A_{n,u_0}(X, \theta) \) with respect to \( \theta \) and let \( A^{(k)}(X, \theta) \) be its probability limit. We assume \( \sup_{(X, \theta) \in X \times \Theta} A^{(k)}(X, \theta) < \infty \) for \( k = 0, 1, 2 \) and \( \inf_{t, \theta} f_{\theta^T X, U}(t, u_0) > 0 \).

C9: Let
\[
\phi_{\theta}(u) = E \left[ \{Y - g(\theta_0^T X)\}^2 \mid U = u \right] \quad (4)
\]
\[
\psi(u) = E \{g'(\theta_0^T X)g_1(X, \theta_0) \otimes X^T \mid U = u \} \quad (5)
\]
\[
\rho(u, x) = E \left[ \{Y - g(\theta_0^T X)\}^2 \mid U = u, X = x \right] \quad (6)
\]
\[
\Delta(u) = E \{\rho(U, X)g_1(X, \theta_0)g_1(X, \theta_0)^T \mid U = u \} \quad (7)
\]
where \( \otimes \) denote the Kronecker product. Here \( \hat{g}^{(k)}(X, \theta) \) is the \( k \)th partial derivative of \( \hat{g}_{u_0}(\theta^T X, \theta) \) with respect to \( \theta \) and \( g_k(X, \theta) \) is its probability limit. We assume \( \phi''_\theta(u) \) is uniformly bounded in \( \theta \). We further assume \( \psi(u), \Delta(u), E(|Y|^3 \mid U = u) \) and the marginal density \( f_U(\cdot) \) of \( U \) are twice differentiable and \( f_U(u_0) > 0 \).

**Lemma 1** Under conditions C1-C8, if \( nh_2 \to \infty \) as \( n \to \infty \) then

\[
\sup_{t,\theta} | A_{n,u_0}(t, \theta) - A(t, \theta) | \xrightarrow{P} 0
\]

where \( A(t, \theta) \) be the probability limit of \( A_{n,u_0}(t, \theta) \).

**PROOF** : Consider the following:

\[
\sup_{t,\theta} | A_{n,u_0}(t, \theta) - A(t, \theta) | \leq \sup_{t,\theta} | A_{n,u_0}(t, \theta) - EA_{n,u_0}(t, \theta) | + \sup_{t,\theta} | EA_{n,u_0}(t, \theta) - A(t, \theta) | \leq I + II \tag{8}
\]

**Analysis of I**

For a suitable sequence \( a_n \to \infty \) we can write

\[
A_{n,u_0}(t, \theta) = (nh_1h_2)^{-1} \sum_{j=1}^{n} Y_j K_1 \left( \frac{\theta^T X_j - t}{h_1} \right) K_2 \left( \frac{U_j - u_0}{h_2} \right) + (nh_1h_2)^{-1} \sum_{j=1}^{n} Y_j I_{\left[ Y_j \in (-a_n,a_n) \right]} K_1 \left( \frac{\theta^T X_j - t}{h_1} \right) K_2 \left( \frac{U_j - u_0}{h_2} \right) = A_{n,1} + A_{n,2}
\]

This implies

\[
\sup_{t,\theta} | A_{n,u_0}(t, \theta) - EA_{n,u_0}(t, \theta) | \leq \sup_{t,\theta} | A_{n,1}(t, \theta) - EA_{n,1}(t, \theta) | + \sup_{t,\theta} | A_{n,2}(t, \theta) - EA_{n,2}(t, \theta) | \leq I_1 + I_2
\]

Consider \( P[I_1 > \epsilon] \). Let \( A_{n,1,j} \) be the \( j \)th summand of \( A_{n,1} \). Then
\[ P[I_1 > \epsilon] = P[\sup_{t,\theta} | (nh_1h_2)^{-1} \sum_{j=1}^n [A_{n,1,j}(t, \theta) - EA_{n,1,j}(t, \theta)] | > \epsilon] \]
\[ = P[\sup_{t,\theta} \sum_{j=1}^n [A_{n,1,j}(t, \theta) - EA_{n,1,j}(t, \theta)] | > \epsilon(nh_1h_2)] \]
\[ = P[\sum_{j=1}^n \sup_{t,\theta} [A_{n,1,j}(t, \theta) - EA_{n,1,j}(t, \theta)] | > \epsilon(nh_1h_2)] \]
\[ \leq \frac{E \sum_{j=1}^n \sup_{t,\theta} [A_{n,1,j}(t, \theta) - EA_{n,1,j}(t, \theta)]}{\epsilon nh_1h_2} \]
\[ \leq \frac{E[\sum_{j=1}^n \sup_{t,\theta} A_{n,1,j}(t, \theta)] + \sup_{t,\theta} |EA_{n,1,j}(t, \theta)|}{\epsilon nh_1h_2} \]
\[ = \frac{nE \sup_{t,\theta} | A_{n,1,j}(t, \theta)|}{\epsilon h_1h_2} + \frac{nE \sup_{t,\theta} | EA_{n,1,j}(t, \theta)|}{\epsilon nh_1h_2} \]
\[ \leq \frac{E \sup_{t,\theta} | A_{n,1,j}(t, \theta)|}{\epsilon h_1h_2} + \frac{\sup_{t,\theta} E | A_{n,1,j}(t, \theta)|}{\epsilon h_1h_2} \]
\[ \leq 2 \frac{E \sup_{t,\theta} | A_{n,1,j}(t, \theta)|}{\epsilon h_1h_2} \]
\[ = 2 \frac{E \sup_{t,\theta} | Y_j I_{[Y_j \notin (-a_n,a_n)]} K_1\left(\frac{\theta^T X_i - t}{h_1}\right) K_2\left(\frac{U_i - u_0}{h_2}\right)|}{\epsilon h_1h_2} \].

Since \( K_1 \) and \( K_2 \) have bounded supports, for suitable constants \( C_1 \) and \( C_2 \) we get
\[ P[I_1 > \epsilon] \leq 2 \frac{C_1C_2}{\epsilon h_1h_2} E |Y_j|^{m}(a_n^{(m-1)}) \]
\[ \leq 2 \frac{C_1C_2}{\epsilon h_1h_2} \left(\frac{E|Y_j|^m}{a_n^m}\right)^{1-1/m} \]
\[ = 2 \frac{C_1C_2}{\epsilon h_1h_2} \frac{E|Y_j|^m}{a_n^{(m-1)}} \].

Therefore \( I_1 \xrightarrow{P} 0 \) uniformly in \((t, \theta)\) if
\[ \epsilon h_1h_2a_n^{(m-1)} \to \infty \). (9)

Consider \( P[I_2 > \epsilon] \). Let \( A_{n,2,j} \) be the \( j \)th summand of \( A_{n,2} \). Then
\[ P[I_2 > \epsilon] = P \left[ \sup_{t,\theta} | (nh_1h_2)^{-1} \sum_{j=1}^n [A_{n,2,j}(t, \theta) - EA_{n,2,j}(t, \theta)] | > \epsilon \right] \]

Let \( \mathcal{X} \) be the set of \( p \)-dimensional vectors and \( \Theta \) be the \( p \)-dimensional parameter space. Without loss of generality assume \( ||x||_\infty \leq 1 \forall x \in \mathcal{X} \). Following Ichimura (1993), partition \( \Theta \)
into $N_1$ cubes with length of a side $(h_1h_2)^\nu\delta$ and $X$ into $N_2$ cubes with length of a side $(h_1h_2)^\nu$, where $\delta$ is a small positive number and $\nu$ is a large constant. Then, $N_1 = \delta^{-p}(h_1h_2)^{-\nu \nu}$ and $N_2 = (h_1h_2)^{-\nu}$ where $p$ is the dimension of the parameter space $\Theta$. Therefore the space $\Theta \times \mathcal{X}$ is partitioned into $N$ cubes with $N = N_1 \times N_2$. Each cube is $p \times p$ dimensional and becomes smaller as $n \to \infty$. Let $B_k^N, k = 1, \ldots, n$ denote all these $p \times p$ cubes. Pick a point $(t_k^N, \theta_k^N)$ from each $B_k^N$. Then

$$P[I_2 > \epsilon] = P \left[ \sup_{(t,\theta) \in B_k^N} (nh_1h_2)^{-1}{\sum_{j=1}^{n} [A_{n,2,j}(t,\theta) - EA_{n,2,j}(t,\theta)]} > \epsilon \right]$$

$$\leq \sum_{k=1}^{N} P \left[ \sup_{(t,\theta) \in B_k^N} (h_1h_2)^{-1}{\sum_{j=1}^{n} [A_{n,2,j}(t,\theta) - EA_{n,2,j}(t,\theta)]} > n\epsilon \right] \tag{10}$$

Each summand of the outside sum of (10) can be decomposed into three parts by adding and subtracting $A_{n,2,j}(t_k^N, \theta_k^N)$ and $EA_{n,2,j}(t_k^N, \theta_k^N)$ within the inside sum. That is

$$P \left[ \sup_{(t,\theta) \in B_k^N} (h_1h_2)^{-1}{\sum_{j=1}^{n} [A_{n,2,j}(t,\theta) - EA_{n,2,j}(t,\theta)]} > n\epsilon \right] \leq P[I_{21k}] + P[I_{22k}] + P[I_{23k}]$$

where

$$P[I_{21k}] = P \left[ (h_1h_2)^{-1}{\sum_{j=1}^{n} [A_{n,2,j}(t_k^N, \theta_k^N) - EA_{n,2,j}(t_k^N, \theta_k^N)]} > \frac{n\epsilon}{2} \right]$$

$$P[I_{22k}] = P \left[ \sup_{(t,\theta) \in B_k^N} (h_1h_2)^{-1}{\sum_{j=1}^{n} [A_{n,2,j}(t,\theta) - A_{n,2,j}(t_k^N, \theta_k^N)]} > \frac{n\epsilon}{4} \right]$$

$$P[I_{23k}] = P \left[ \sup_{(t,\theta) \in B_k^N} (h_1h_2)^{-1}{\sum_{j=1}^{n} [EA_{n,2,j}(t_k^N, \theta_k^N) - EA_{n,2,j}(t,\theta)]} > \frac{n\epsilon}{4} \right] .$$

Then

$$P[I_2 > \epsilon] \leq \sum_{k=1}^{N} P[I_{21k}] + \sum_{k=1}^{N} P[I_{22k}] + \sum_{k=1}^{N} P[I_{23k}] = I_1^1 + I_2^2 + I_2^3$$

Now we will show that these 3 terms converge in probability to 0 as $n \to \infty$. Consider $I_2^1$.

$$P[I_{21k}] = P \left[ \sum_{j=1}^{n} W_{jn} > \frac{n\epsilon h_1h_2}{2} \right]$$

where

$$W_{jn} = [A_{n,2,j}(t_k^N, \theta_k^N) - EA_{n,2,j}(t_k^N, \theta_k^N)] .$$
Using the definition of \( A_{n,2} \) and boundedness of the two kernels, we have
\[
|W_{in}| \leq 2a_nC_1C_2
\]
\[
\var\{W_{jn}\} \leq E\{A_{n,2,j}^2\}
\]
\[
\leq E \left[ |Y|^2 K_1^2 \left( \frac{\theta^T X - t_N^N}{h_1} \right) K_2^2 \left( \frac{u - u_0}{h_2} \right) \right]
\]
\[
= \int \int \psi(s, u) K_1^2 \left( \frac{s - t_N^N}{h_1} \right) K_2^2 \left( \frac{u - u_0}{h_2} \right) f_{\theta^T X, U}(s, u) dsdu
\]
where \( \psi(s, u) = E \left[ |Y|^2 \mid \theta^T X = s, U = u \right] \). Assuming \( \psi(\cdot, \cdot) \) and \( f_{\theta^T X, U}(\cdot, \cdot) \) are differentiable with bounded derivatives, standard arguments yield \( \var\{W_{jn}\} = O(h_1h_2) \). Applying Bernstein’s inequality to \( P[I_{21k}] \) we get \( P[I_{21k}] \leq \exp(-d_{n1}/2) \) where
\[
d_{n1} = \frac{n\epsilon^2h_1h_2C}{1 + a_n\epsilon C}.
\]
Here and what follows we use \( C \) to denote a generic positive constant. Therefore
\[
\sum_{k=1}^N P[I_{21k}] \leq N \exp(-d_{n1})
\]
\[
= [\delta^{-p}(h_1h_2)^{-2\nu t}] \exp(-d_{n1})
\]
\[
= \delta^{2p} \exp \left[ -d_{n1} \left\{ 1 + 4\nu \frac{\ln(h_1h_2)}{d_{n1}} \right\} \right]
\]
Note that \(-\ln(h_1h_2)/d_{n1} \to 0 \Rightarrow d_{n1} \to \infty \) faster than \(-\ln(h_1h_2)\). Therefore if
\[
-\ln(h_1h_2)/d_{n1} \to 0
\]
then \( \sum_{k=1}^N P[I_{21k}] \to 0 \) as \( n \to \infty \) which implies that \( I_{12} = O(c_{1n}) \) where
\[
c_{1n} = \exp \left[ -d_{n1} \left\{ 1 + C \frac{\ln(h_1h_2)}{d_{n1}} \right\} \right] .
\]
Now consider \( I_{22} \). Note that
\[
P[I_{22k}] \leq P \left[ (h_1h_2)^{-1} \sum_{j=1}^n \sup_{(t, \theta) \in B_k^N} \left| A_{n,2,j}(t, \theta) - A_{n,2,j}(t^N_k, \theta^N_k) \right| > \frac{n\epsilon}{4} \right].
\]
Denoting
\[
T_{nk} = \sum_{j=1}^n \sup_{(t, \theta) \in B_k^N} \left| A_{n,2,j}(t, \theta) - A_{n,2,j}(t^N_k, \theta^N_k) \right|
\]
we get
\[
P[I_{22k}] \leq \left[ |T_{nk} - ET_{nk} + ET_{nk}| > \frac{n\epsilon(h_1h_2)}{4} \right]
\]
\[
= T_{n1} + T_{n2}
\]
where
\[ T_{nk1} = P \left[ |T_{nk} - ET_{nk}| > \frac{n\epsilon(h_1h_2)}{8} \right] \]
and
\[ T_{nk2} = P \left[ ET_{nk} > \frac{n\epsilon(h_1h_2)}{8} \right]. \]

Before we analyze these two terms, we establish certain bounds related to \( T_{nk}. \) First note that
\[
\sup_{(t,\theta)\in B_k^N} |A_{n,2,j}(t, \theta) - A_{n,2,j}(t_k^N, \theta_k^N)|
= \sup_{(t,\theta)\in B_k^N} Y_j I_{[y_j \in (-a_n, a_n))] } K_2 \left( \frac{U_j - u_0}{h_2} \right) \left\{ K_1 \left( \frac{\theta^T X_j - t}{h_1} \right) - K_1 \left( \frac{(\theta_k^N)^T X_j - t_k^N}{h_1} \right) \right\}
\]

Assuming \( K_1(\cdot) \) is differentiable, for a suitable mean value \( \bar{c} \), we get
\[
\sup_{(t,\theta)\in B_k^N} \left| K_1 \left( \frac{\theta^T X_j - t}{h_1} \right) - K_1 \left( \frac{(\theta_k^N)^T X_j - t_k^N}{h_1} \right) \right|
= \sup_{(t,\theta)\in B_k^N} \left| \frac{(\theta^T X_j - t) - \{(\theta_k^N)^T X_j - t_k^N\}}{h_1} \right| K_1' (\bar{c})
\leq \frac{C}{h_1} \sup_{(t,\theta)\in B_k^N} \left| (\theta^T X_j - t) - \{(\theta_k^N)^T X_j - t_k\} \right| .
\]

Note that \( t_k^N \) and \( t \) are points from \( B_k^N \) and therefore, \( (t_k^N - t) \) can be written as \( (\theta_k^N)^T x^* \) for some suitable \( x^* \in \mathcal{X} \). Also note that \( ||\theta - \theta_k^N||_2 \leq \sqrt{p(h_1h_2)^{2\nu}\delta^2} = C(h_1h_2)^\nu \). Then from the boundedness of the \( X \) vectors and the compactness of the parameter space \( \Theta \), we get
\[
\sup_{(t,\theta)\in B_k^N} |A_{n,2,j}(t, \theta) - A_{n,2,j}(t_k^N, \theta_k^N)| \leq \frac{C a_n(h_1h_2)^\nu}{h_1} . \tag{12}
\]

Now consider the expectation of the left hand side of (12).
\[
E \left[ \sup_{(t,\theta)\in B_k^N} |A_{n,2,j}(t, \theta) - A_{n,2,j}(t_k^N, \theta_k^N)| \right] \leq \frac{(h_1h_2)^\nu}{h_1} E \left[ \sup_{(t,\theta)\in B_k^N} Y_j I_{[y_j \in (-a_n, a_n))] } K_2 \left( \frac{U_j - u_0}{h_2} \right) \right]
\]
\[
= \frac{(h_1h_2)^\nu}{h_1} \int \phi_1(u) K_2 \left( \frac{u - u_0}{h_2} \right) f_U(u) du
\]
where \( \phi_1(u) = E[|Y| \mid U = u] \) and \( f_U(u) \) is the density of \( U \). Assuming \( \phi \) and \( f_U \) are differentiable with bounded derivatives, standard arguments yield
\[
E \left[ \sup_{(t,\theta)\in B_k^N} |A_{n,2,j}(t, \theta) - A_{n,2,j}(t_k^N, \theta_k^N)| \right] = O \left( h_1^{\nu-1} h_2^{\nu+1} \right) . \tag{13}
\]
Now consider $T_{nk1}$. To apply Bernstein’s inequality, let $W_{jn} = T_{nk1} - ET_{nk1}$. Using (12) and (13) we get $|W_{jn}| \leq |T_{nk1}| + |ET_{nk1}| \leq Ca_n h_1^{\nu-2}h_2^{\nu-1}$. Also

$$Var[W_{jn}] = Var[T_{nk1}]$$

$$\leq E[T_{nk1}^2]$$

$$\leq \frac{C(h_1h_2)^{2\nu}}{h_1^2} E \left[ Y^2 K_2^2 \left( \frac{U - u_0}{h_2} \right) \right]$$

$$= \frac{C(h_1h_2)^{2\nu}}{h_1^2} \int \phi_2(u) K_2^2 \left( \frac{u - u_0}{h_2} \right) f_U(u)\,du$$

where $\phi_2(u) = E[Y^2 | U = u]$. As before, assuming $\phi_2$ and $f_U$ is differentiable with bounded derivatives, we get $Var[W_{jn}] = O(h_1^{2\nu-4}h_2^{2\nu-1})$. Now applying Bernstein’s inequality to $T_{nk1}$ we get $T_{nk1} \leq \exp(-d_{n2}/2)$ where

$$d_{n2} = \frac{n\epsilon^2 h_1 h_2 C}{h_1^{2\nu-5} h_2^{2\nu-4} + \epsilon a_n h_1^{\nu-2} h_2^{\nu-1} C}.$$ 

If $\nu \geq 3$ then $d_{n2} > d_{n1}$ for large enough $n$. If (11) is satisfied then $d_{n1} \to \infty \Rightarrow d_{n2} \to \infty$ as $n \to \infty$ which implies

$$\sum_{k=1}^{N} T_{nk1} \leq N \exp(-d_{n2}).$$

Therefore if (11) is satisfied then $\sum_{k=1}^{N} T_{nk1} \to 0$ as $n \to \infty$.

Now consider $T_{nk2}$.

$$T_{nk2} = P \left[ ET_{nk} > \frac{n\epsilon(h_1h_2)}{8} \right]$$

$$= P \left[ n E \sup_{(t, \theta) \in B_k^N} |A_{n,2,j}(t, \theta) - A_{n,2,j}(t_k^N, \theta_k^N)| > \frac{n\epsilon(h_1h_2)}{8} \right]$$

$$\leq P \left[ O(h_1^{\nu-1}h_2^{\nu+1}) > \frac{c(h_1h_2)}{8} \right]$$

where the last inequality follows from (13). For $\nu \geq 2$ we get $T_{nk2} \to 0$ as $n \to \infty$ which implies $\sum_{k=1}^{N} T_{nk2} \to 0$. Therefore we have $\sum_{k=1}^{N} P[I_{2k}] \to 0$ which yields $I_2^2 \to 0$ as $n \to \infty$. Moreover we can choose $\nu \geq 3$ so that the rate is faster than $c_{1n}$ because for $\nu \geq 3$, $d_{n2} > d_{n1}$.

**Remark 1** $N$ is the number of cubes and is fixed for all $n.$
Now consider $I_2^3$.

\[
P[I_{23k}] \leq P \left[ \sup_{(t, \theta) \in B_k^n} \left| (h_1 h_2)^{-1} \sum_{j=1}^{n} A_{n, 2, j}(t_k^n, \theta_k^n) - A_{n, 2, j}(t, \theta) \right| > \frac{n \epsilon}{4} \right]
\]

\[
\leq P \left[ \sup_{(t, \theta) \in B_k^n} \left| A_{n, 2, j}(t_k^n, \theta_k^n) - A_{n, 2, j}(t, \theta) \right| > \frac{n \epsilon (h_1 h_2)}{4} \right]
\]

\[
= P \left[ n \sup_{(t, \theta) \in B_k^n} \left| A_{n, 2, 1}(t_k^n, \theta_k^n) - A_{n, 2, 1}(t, \theta) \right| > \frac{n \epsilon (h_1 h_2)}{4} \right]
\]

\[
\leq P \left[ O \left( h_1^{\nu-1} h_2^{\nu+1} \right) > \frac{\epsilon (h_1 h_2)}{4} \right]
\]

where the last inequality follows from (13). As before, for $\nu \geq 2$ we have $\sum_{k=1}^{N} P[I_{23k}] \to 0$ which yields $I_2^3 \to 0$ as $n \to \infty$ as desired. Therefore if (11) is satisfied we have $I \overset{P}{\to} 0$ as $n \to \infty$.

**Analysis of II**

Now consider the second term of (8).

\[
E[A_n(t, \theta)] = E \left[ \frac{1}{nh_1 h_2} \sum_{j=1}^{n} Y_j K_1 \left( \frac{\theta^T X_j - t}{h_1} \right) K_2 \left( \frac{U_j - u_0}{h_2} \right) \right]
\]

\[
= \frac{1}{h_1 h_2} \int \int \psi(s, u) K_1 \left( \frac{s - t}{h_1} \right) K_2 \left( \frac{u - u_0}{h_2} \right) f_{\theta^T X, U}(s, u) ds du
\]

where $\psi(s, u) = E[Y \mid \theta^T X = s, U = u]$ and $f_{\theta^T X, U}(s, u)$ is the joint density of $\theta^T X, U$. Let $\psi^* = \psi f$ and using a change of variables we can rewrite the above expression as

\[
\frac{1}{h_1 h_2} \int \int \psi^*(t + h_1 v, u_0 + h_2 w) K_1 (v) K_2 (w) h_1 h_2 dv dw.
\]

Assuming $\psi^*$ is differentiable and uniformly bounded in $(t, \theta)$, a Taylor series expansion yields

\[
E[A_n(t, \theta)] = \psi^*(t, u_0) + O(h_1^2) + O(h_2^2) + O(h_1 h_2).
\]

Denoting $\psi^*(t, u_0) = A(t, \theta)$ we get

\[
E[A_n(t, \theta)] - A(t, \theta) = O(h_1^2) + O(h_2^2) + O(h_1 h_2).
\]

Therefore term II of (8) converge to zero in probability as $n \to \infty$. 

10
Uniform convergence

We will now show the uniform convergence of \( \hat{g}_{u_0}(t, \theta) = A_{n,u_0}(t, \theta)/B_{n,u_0}(t, \theta) \) in \( t, \theta \). We have shown that
\[
\sup_{t, \theta} | A_{n,u_0}(t, \theta) - A(t, \theta) | \xrightarrow{p} 0
\]
as \( n \to \infty \). Let the rate of convergence be \( a_{n_0} \). Similarly, we can show that
\[
\sup_{t, \theta} | B_{n,u_0}(t, \theta) - B(t, \theta) | \xrightarrow{p} 0
\]
as \( n \to \infty \) with convergence rate \( b_{n_0} \). To avoid \( \inf_{t, \theta} |B_{n,u_0}(t, \theta)| = 0 \) in technical arguments we will add a \( c_n \) to \( B_{n,u_0} \) and pick \( c_n \) appropriately so that the estimator is uniformly convergent. Note that
\[
A(t, \theta) = E[Y|\theta^TX = t, U = u_0]f_{\theta^TX,U}(t, u_0)
\]
and
\[
B(t, \theta) = f_{\theta^TX,U}(t, u_0) .
\]
We will assume \( \inf_{t, \theta} |B(t, \theta)| > 0 \). We now show that
\[
\sup_{t, \theta} \left| \frac{A_{n}(t, \theta)}{B_{n}(t, \theta)} - \frac{A(t, \theta)}{B(t, \theta)} \right| \xrightarrow{p} 0
\]
as \( n \to \infty \). Suppressing the argument \( (t, \theta) \), we can write
\[
\left| \frac{A_{n,u_0} - A}{B_{n,u_0} - B} \right| = \left| \frac{A_{n,u_0} - A}{B_{n,u_0}} \right| + \left| \frac{A}{B_{n,u_0} - B} \right| \leq \frac{|A_{n,u_0} - A|}{|B_{n,u_0}|} + \frac{|A||B - B_{n,u_0}|}{|B_{n,u_0}B|} \leq \frac{\sup_{t, \theta} |A_{n,u_0}(t, \theta) - A(t, \theta)|}{\inf_{t, \theta} |B_{n,u_0}(t, \theta)|} + \frac{\sup_{t, \theta} |A(t, \theta)||B(t, \theta) - B_{n,u_0}(t, \theta)|}{\inf_{t, \theta} |B_{n,u_0}(t, \theta)B(t, \theta)|} .
\]
Note that \( \inf_{t, \theta} |B_{n,u_0}(t, \theta)| = c_n \) and using condition C8 we have
\[
\sup_{t, \theta} \left| \frac{A_{n,u_0}(t, \theta)}{B_{n,u_0}(t, \theta)} - \frac{A(t, \theta)}{B(t, \theta)} \right| = O \left( a_{n_0}/c_n \right) + O \left( b_{n_0}/c_n \right) .
\]
This implies
\[
\sup_{t, \theta} | \hat{g}_{u_0}(t, \theta) - g_0(t, \theta) | = o_p(1)
\]
where \( g_0(t, \theta) = A(t, \theta)/B(t, \theta) \). This completes the proof.

Next we will establish two rate results pertaining to the uniform convergence of the first and the second partial derivatives of the link function estimator with respect to \( \theta \). Let \( m \) denote the highest absolute moment of the response \( Y \). Recall that \( A^{(1)}(X, \theta) \) and \( A^{(2)}(X, \theta) \) be the respective probability limits of the first and the second partial derivatives of \( A_{n,u_0}(\theta^TX, \theta) \) with respective to \( \theta \).
Lemma 2 Under conditions C1-C8,
\[
\sup_{(X,\theta)\in(X\times\Theta)} \left| \frac{\partial A_n(\theta^T X, \theta)}{\partial \theta} - A(2)(X,\theta) \right| = O_p(a_{n1})
\]

Lemma 3 Under conditions C1-C8,
\[
\sup_{(X,\theta)\in(X\times\Theta)} \left| \frac{\partial^2 A_n(\theta^T X, \theta)}{\partial \theta \partial \theta^T} - A(2)(X,\theta) \right| = O_p(a_{n2})
\]

The proofs of these two lemmas follow the same lines of arguments and hence only the proof of Lemma 3 will be given.

PROOF : As in Lemma 1, we can decompose
\[
\left| \frac{\partial^2 A_n(\theta^T X, \theta)}{\partial \theta \partial \theta^T} - A(2)(X,\theta) \right|
\]
into two terms as
\[
\left| \frac{\partial^2 A_n(\theta^T X, \theta)}{\partial \theta \partial \theta^T} - A(2)(X,\theta) \right| \leq \left| \frac{\partial^2 A_n(\theta^T X, \theta)}{\partial \theta \partial \theta^T} - E \frac{\partial^2 A_n(\theta^T X, \theta)}{\partial \theta \partial \theta^T} - A(2)(X,\theta) \right| + \left| E \frac{\partial^2 A_n(\theta^T X, \theta)}{\partial \theta \partial \theta^T} - A(2)(X,\theta) \right| = I + II.
\]
Term I can be analyzed using the techniques given in Lemma 1 term I with an additional \(h_2^2\) in the denominators of the order terms. Therefore we get \(I = a_{n2,1}\) where
\[
a_{n2,1} = O \left\{ \frac{1}{h_1^3 h_2 a_n (m-1)} \right\} + O \left[ \exp \left\{ -d''_{n1} \left( 1 + \frac{lnh_1 h_2}{a_n h_1^2} \right) \right\} \right]
\]
where
\[
d''_{n1} = \frac{Cnh_1^5 h_2}{1 + Ca_n h_1^2}
\]
and \(a_n\) is a sequence that satisfy \(a_n \to \infty\) as \(n \to \infty\) whose rate will be determined later.

Now consider the term II.
\[
\frac{\partial^2 A_n(\theta^T X, \theta)}{\partial \theta \partial \theta^T} = \frac{1}{nh_1^3 h_2} \sum_{j=1}^{n} Y_j K'' \left( \frac{\theta^T X_j - \theta^T X}{h_1} \right) K_2 \left( \frac{U_j - u_0}{h_2} \right) (X_j - X)(X_j - X)^T.
\]
Note that this is a \(p \times p\) matrix and we will analyze only the \((1,1)\) element. First we need to compute the expectation of this element.
\[
E \left[ \left\{ \frac{\partial^2 A_n(\theta^T X, \theta)}{\partial \theta \partial \theta^T} \right\}_{(1,1)} \right] = E \left[ E \left( \cdot \mid \theta^T X \right) \right] = \int \phi'(s)f_{\theta^T X}(s)ds
\]
where

$$
\phi^*_2(t) = E \left[ \frac{1}{h_1 h_2} Y_j K''_1 \left( \frac{\theta^T X_j - \theta^T X}{h_1} \right) K_2 \left( \frac{u - u_0}{h_2} \right) [(X_j - X_1)(X_j - X)^T]_{1,1} \mid \theta^T X = t \right] 
$$

$$
= E[E(\cdot \mid \theta^T X_j = s, U = u, \theta^T X = t) \mid \theta^T X = t]
$$

$$
= \frac{1}{h_1^3 h_2} \int \int \psi(s, u, t) K_1 \left( \frac{s - t}{h_1} \right) K_2 \left( \frac{u - u_0}{h_2} \right) f_{\theta^T X_1, U \mid \theta^T X = t}(s, u|t) ds du
$$

$$
= \frac{1}{h_1^3 h_2} \int \int \psi^*(t + h_1 v, u_0 + h_2 w, t) K''_1(v) K_2(w) h_1 dv h_2 dw
$$

$$
= T_1 + \ldots + T_3 + \frac{1}{2}(T_4 + \ldots + T_6) + \frac{1}{6} \sum_{k=1}^9 T_{7,k}
$$

where $$\psi^* = \psi_{\theta^T X_1, U \mid \theta^T X = t}$$ and

$$
\psi(s, u, t) = E \left[ Y \left\{ (X_j - X)_{1,1} \right\}^2 \mid \theta^T X_j = s, U = u, \theta^T X = t \right].
$$

We will analyze these terms next. We need further notation and conditions to analyze these terms. Let $$\phi^*_{x/y}(\cdot, \cdot, \cdot) = \frac{\partial^{j+k} \phi^*_1(x, y, t)}{\partial x^j \partial y^k}$$. Using our assumptions on the two kernels in condition C7, we see that all odd moments of $$K_1$$ and $$K_2$$ are zero and all even moments are non zero. It is also easy to see that all odd moments of $$K''_1$$ and $$K''_2$$ are zero and all even moments are non zero. In addition to these we require $$\int K''_1(s) ds = 0$$. These conditions are satisfied by most kernel that are used in practice. For example $$K(x) = C(1 - x^2)^2 I_{[-1,1]}(x)$$ is one such kernel. Under these conditions we have

$$
T_1 = \frac{1}{h_1^2} \phi^*_x(t, u_0, t) \int K''_1(v) dv \int K_2(w) dw = 0
$$

$$
T_2 = \frac{1}{h_1^2} \phi^*_x(t, u_0, t) \int \int h_1 v K''_1(v) K_2(w) dv dw = 0
$$

$$
T_3 = \frac{1}{h_1^2} \phi^*_y(t, u_0, t) \int \int K''_1(v) h_2 w K_2(w) dv dw = 0
$$

$$
T_4 = \frac{1}{h_1^2} \phi^*_{x^2}(t, u_0, t) \int \int h_1^2 v^2 K''_1(v) K_2(w) dv dw = \phi^*_{xx}(t, u_0, t)
$$

$$
T_5 = \frac{1}{h_1^2} \phi^*_{y^2}(t, u_0, t) \int \int K''_1(v) h_2^2 w^2 K_2(w) dv dw = 0
$$

$$
T_6 = \frac{1}{h_1^2} \phi^*_{xy}(t, u_0, t) \int \int h_1 v K''_1(v) h_2 w K_2(w) dv dw = 0.
$$
Therefore we get
\[\phi_T \phi_T\] on compact set. So if we assume

Now we will analyze \(T_7\).

\[
T_{7,1} = \frac{1}{h_1^2} \phi^* \phi^* (t, u_0, t) \int \int h_1^3 v^3 K'_1(v) K_2(w) dv dw = 0
\]

\[
T_{7,2} = \frac{1}{h_1^2} \phi^* \phi^* (t, u_0, t) \int \int K''_1(v) h_2^2 w^2 K_2(w) dv dw = 0
\]

\[
T_{7,3} = \frac{1}{h_1^2} \phi^* \phi^* (t, u_0, t) \int \int h_1^2 v^2 K''_1(v) h_2 w K_2(w) dv dw = 0
\]

\[
T_{7,4} = \frac{1}{h_1^2} \phi^* \phi^* (t, u_0, t) \int \int h_1 v K''_1(v) h_2^2 w^2 K_2(w) dv dw = 0
\]

\[
T_{7,5} = \frac{1}{h_1^2} \int \int \phi^* (\bar{c}_1, \bar{c}_2, t) h_1^4 v^4 K''_1(v) K_2(w) dv dw
\]

\[
T_{7,6} = \frac{1}{h_1^2} \int \int \phi^* (\bar{c}_3, \bar{c}_4, t) K''_1(v) h_2^4 w^4 K_2(w) dv dw
\]

\[
T_{7,7} = \frac{1}{h_1^2} \int \int \phi^* (\bar{c}_5, \bar{c}_6, t) h_1^3 v^3 K'_1(v) h_2 w K_2(w) dv dw
\]

\[
T_{7,8} = \frac{1}{h_1^2} \int \int \phi^* (\bar{c}_7, \bar{c}_8, t) h_1^2 v^2 K'_1(v) h_2^2 w^2 K_2(w) dv dw
\]

\[
T_{7,9} = \frac{1}{h_1^2} \int \int \phi^* (\bar{c}_9, \bar{c}_{10}, t) h_1 v K'_1(v) h_2^3 w^3 K_2(w) dv dw
\]

where \(\bar{c}_1, \ldots, \bar{c}_{10}\) are the corresponding in between values of the Taylors expansion of \(\phi^*\). Note that \(\|X\|_{\infty} \leq 1\) and \(\theta\) is in a compact set and hence \(\theta^T X\) will be in a bounded interval. Also recall that \(U \in [0, 1]\). Therefore \(\phi^* (s, u, t) = \psi (s, u, t) f_{\theta^T X} (s, u | t)\) is defined on compact set. So if we assume \(\phi (s, u, t)\) is continuous in \(t\), then all the order terms in \(T_{7,5}, \ldots, T_{7,9}\) are free of \(t\) and their magnitudes are listed below.

\[
T_{7,5} = O(h_1^2)
\]

\[
T_{7,6} = O(h_1^{-2} h_2^4)
\]

\[
T_{7,7} = O(h_1 h_2)
\]

\[
T_{7,8} = O(h_2^2)
\]

\[
T_{7,9} = O(h_1^{-1} h_2^3)
\]

Therefore we get \(\phi^*_2 (t) = \phi^*_2 (t, u_0, t) + \sum_{k=5}^{9} T_{7,k}\) and hence

\[
E \left[ \frac{\partial^2 A_n (\theta^T X, \theta)}{\partial \theta \partial \theta^t} \right]_{(1,1)} = \int \phi^*_2 (s, u_0, s) f_{\theta^T X} (s) ds + \sum_{k=5}^{9} T_{7,k}
\]
which yields \( II = a_{n,2,2} \) where
\[
a_{n,2,2} = O(h_1^2) + O(h_2^2) + O(h_1 h_2) + O(h_1^{-1} h_2^3) + O(h_1^{-2} h_2^4).
\]
Finally we get the convergence rate \( a_{n,2} = a_{n,2,1} + a_{n,2,2} \) which completes the proof.

**Uniform Convergence**

Now we will derive the conditions for uniform convergence of the second partial derivative of the link function estimator. Recall that \( \hat{g}_{u_0}(t, \theta) = A_{n,u_0}(t, \theta)/B_{n,u_0}(t, \theta) \) and therefore
\[
\frac{\partial \hat{g}_{u_0}(\theta T X, \theta)}{\partial \theta} = \frac{A_n^{(1)}(\theta T X, \theta)}{B_n(\theta T X, \theta)} - \frac{A_n(\theta T X, \theta)B_n^{(1)}(\theta T X, \theta)}{B_n^2(\theta T X, \theta)}.
\]

We will suppress the argument \( \theta T X, \theta \) and the subscript \( u_0 \) of these functions for ease of exposition. Then we have
\[
\frac{\partial^2 \hat{g}_{u_0}(\theta T X, \theta)}{\partial \theta \partial \theta^T} = \frac{A_n^{(2)}}{B_n} - \frac{A_nB_n^{(1)}}{B_n^2} - \frac{B_n^{(1)}\{A_n^{(1)}\}^T}{B_n^2} + 2\frac{A_nB_n^{(1)}\{B_n^{(1)}\}^T}{B_n^3}.
\]

Let \( g_2(X, \theta) \) be the probability limit of \( \frac{\partial^2 \hat{g}_{u_0}(\theta T X, \theta)}{\partial \theta \partial \theta^T} \). Now we will show that
\[
\sup_{X, \theta} \left| \frac{\partial^2 \hat{g}_{u_0}(\theta T X, \theta)}{\partial \theta \partial \theta^T} - g_2(X, \theta) \right| \overset{p}{\longrightarrow} 0.
\]

We will show the proof for a \((i, j)\) element in the above matrices. Consider the elementwise difference.
\[
\left| \left\{ \frac{\partial^2 \hat{g}_{u_0}(\theta T X, \theta)}{\partial \theta \partial \theta^T} \right\}_{(i,j)} - \{g_2(X, \theta)\}_{(i,j)} \right| \leq |I| + \ldots + |V|.
\]

From here on we will suppress the subscript \((i, j)\) and analyze these terms. Recall that \( A_1(X, \theta) \) and \( A_2(X, \theta) \) be the respective probability limits of the first and the second partial derivatives of \( A_n(\theta T X, \theta) \) with respective to \( \theta \). Let \( B_n^{(k)}(X, \theta) \) and \( B^{(k)}(X, \theta) \) are the analogous \( k \)th partial derivative and probability limit involving \( B_{n,u_0}(X, \theta) \).
\(|I| = \left| \frac{A_n^{(2)}}{B_n} - \frac{A^{(2)}}{B} \right|

= \left| \frac{A_n^{(2)}}{B_n} - \frac{A^{(2)}}{B_n} + \frac{A^{(2)}}{B_n} - \frac{A^{(2)}}{B} \right|

\leq \frac{|A_n^{(2)} - A^{(2)}|}{B_n} + \frac{|A^{(2)}| |B - B_n|}{|B_nB|}

\leq \frac{|A_n^{(2)} - A^{(2)}|}{c_n} + \frac{C_A^{(2)} |B - B_n|}{cc_n}

where \(c_n\) and \(c\) are defined in Lemma 1 and \(C_A^{(2)} = \sup_{X,\theta} |A^{(2)}(\theta^T X, \theta)|\). Therefore \(\sup_{X,\theta} |I| \to 0\) if

\[
a_{n2} \frac{c_n}{c} \to 0 \\
b_{n0} \frac{c_n}{c} \to 0.
\]

Now consider \(|II|\).

\(|II| = \left| \frac{A^{(1)} \{ B^{(1)} \}^T}{B^2} - \frac{A_n^{(1)} \{ B_n^{(1)} \}^T}{B_n^2} \right|

= \left| \frac{A^{(1)} \{ B^{(1)} \}^T}{B^2} - \frac{A^{(1)} \{ B^{(1)} \}^T}{B_n^2} + \frac{A^{(1)} \{ B^{(1)} \}^T}{B^2} - \frac{A_n^{(1)} \{ B_n^{(1)} \}^T}{B_n^2} \right|

\leq \frac{|A^{(1)} \{ B^{(1)} \}^T|}{B_n^2 B^2} |B_n^2 - B^2| + \frac{|A^{(1)} \{ B^{(1)} \}^T - A_n^{(1)} \{ B_n^{(1)} \}^T|}{B_n^2}

\leq \frac{C_{AB} |B_n^2 - B^2|}{c^2 c_n^2} + \frac{|A^{(1)} \{ B^{(1)} \}^T - A_n^{(1)} \{ B^{(1)} \}^T + A_n^{(1)} \{ B^{(1)} \}^T - A_n^{(1)} \{ B_n^{(1)} \}^T|}{c_n^2}

\leq II_1 + II_2 + II_3
where $C^{(1)}_{AB}$ is a constant such that $\sup_{X, \theta} \left| A^{(1)}(\theta^TX, \theta) \left\{ B^{(1)}(\theta^TX, \theta) \right\}^T \right| \leq C^{(1)}_{AB}$ and

\[
II_1 = \frac{C^{(1)}_{AB} \left| B_n^2 - B^2 \right|}{c^2 c_n^2},
\]
\[
II_2 = \frac{\left| \left( A^{(1)} - A_n^{(1)} \right) \left\{ B^{(1)} \right\}^T \right|}{c_n^2},
\]
\[
II_3 = \frac{\left| A_n^{(1)} \left\{ B^{(1)} \right\}^T - \left\{ B_n^{(1)} \right\}^T \right|}{c_n^2}.
\]

$II_1 \overset{p}{\to} 0$ if $\frac{b_n^2}{c_n^2} \to 0$. (16)

Now consider $II_2$.

\[
II_2 \leq \frac{C^{(1)}_B \left| A^{(1)} - A_n^{(1)} \right|}{c_n^2},
\]

where $\sup_{X, \theta} \left| B^{(1)}(\theta^TX, \theta) \right| \leq C^{(1)}_B$. Therefore $II_2 \overset{p}{\to} 0$ if $\frac{a_n}{c_n^2} \to 0$. (17)

Now consider $II_3$.

\[
II_3 \leq \frac{\left\{ A_n^{(1)} - A^{(1)} \right\} \left\{ B^{(1)} \right\}^T - \left\{ B_n^{(1)} \right\}^T \right|}{c_n^2} + \frac{\left| A^{(1)} \left\{ B^{(1)} \right\}^T - \left\{ B_n^{(1)} \right\}^T \right|}{c_n^2}
\]

\[
\leq \frac{\left\{ A_n^{(1)} - A^{(1)} \right\} \left\{ B^{(1)} \right\}^T - \left\{ B_n^{(1)} \right\}^T \right|}{c_n^2} + \frac{C^{(1)}_A \left\{ B^{(1)} \right\}^T - \left\{ B_n^{(1)} \right\}^T \right|}{c_n^2}
\]

where $\sup_{X, \theta} \left| A^{(1)}(\theta^TX, \theta) \right| \leq C^{(1)}_A$. Therefore $II_3 \overset{p}{\to} 0$ if

\[
\frac{a_n b_n}{c_n} \to 0.
\]

(18)

(19)

Now consider $III$.

\[
III \leq \frac{C^{(2)}_A \left| B_n^2 - B^2 \right|}{c^2 c_n^2} + \frac{C^{(2)}_B \left| B^{(2)} - B_n^{(2)} \right|}{c_n^2} + \frac{|A - A_n| \left| B_n^{(2)} - B^{(2)} \right|}{c_n^2} + \frac{C^{(2)}_B |A - A_n|}{c_n^2}
\]
and therefore $|III| \overset{p}{\to} 0$ if

\[
\frac{b_{n0}^2}{c_n^2} \to 0 \quad (20)
\]

\[
\frac{b_{n2}}{c_n} \to 0 \quad (21)
\]

\[
\frac{b_{n2}a_{n0}}{c_n} \to 0 \quad (22)
\]

\[
\frac{a_{n0}}{c_n^2} \to 0 \quad (23)
\]

Now consider IV. Note that it is the transpose of term 'II' and hence we skip this term. Consider V.

\[
|V| \leq \frac{2|A_n\begin{bmatrix} B_n^{(1)} \{ B_n^{(1)} \}^T \end{bmatrix} - A \begin{bmatrix} B^{(1)} \{ B^{(1)} \}^T \end{bmatrix}|}{|B_n^3|} + \frac{2|A\begin{bmatrix} B^{(1)} \{ B^{(1)} \}^T \end{bmatrix}| |B^n_3 - B^n_3|}{|B^n_3 B^n_3|} \\
\leq \frac{2 \left| B_n^{(1)} \begin{bmatrix} B_n^{(1)} \{ B_n^{(1)} \}^T \end{bmatrix} \right| |A_n - A|}{c_n^3} + \frac{2C_A \left| B_n^{(1)} \begin{bmatrix} B_n^{(1)} \{ B_n^{(1)} \}^T - B^{(1)} \{ B^{(1)} \}^T \end{bmatrix} \right|}{c_n^3} + \frac{2C_A \left| B^{(1)} \begin{bmatrix} B^{(1)} \{ B^{(1)} \}^T \end{bmatrix} \right|}{c_n^3} + \frac{2|A_n - A| |B^{(1)} \{ B^{(1)} \}^T|}{c_n^3} + \frac{C_A |B^{(1)} \{ B^{(1)} \}^T|}{c_n^3} + \frac{C_AC_{AB} |B^n_3 - B^n_3|}{c_n^3 c_n^3}
\]

and therefore $|V| \overset{p}{\to} 0$ if

\[
\frac{a_{n0}b_{n1}^2}{c_n^3} \to 0 \quad (24)
\]

\[
\frac{a_{n0}}{c_n} \to 0 \quad (25)
\]

\[
\frac{b_{n1}^2}{c_n} \to 0 \quad (26)
\]

\[
\frac{b_{n0}^3}{c_n^2} \to 0 \quad (27)
\]

This completes the proof. □
3 Consistency of Coefficient Estimator

We will show that \( \hat{\theta} \) is a consistent estimator of \( \theta_0 = \{\beta_1(u_0), \ldots, \beta_p(u_0)\}^T \).

**THEOREM 1** Under assumptions C1-C9, as \( n \to \infty \) if \( nh_2 \to \infty \) then \( \hat{\theta} \) is a consistent estimator of \( \theta_0 \).

**PROOF :** By definition of \( \hat{\theta} \), \( P \left[ M_n^{1/2}(\hat{\theta}) \leq M_n^{1/2}(\theta_0) \right] = 1 \). Let \( B_r(\theta_0) \) denote the open ball of radius \( r > 0 \) centered at \( \theta_0 \). Let \( A \) be the event \( M_n^{1/2}(\hat{\theta}) \leq M_n^{1/2}(\theta_0) \). Then

\[
P \left[ M_n^{1/2}(\hat{\theta}) \leq M_n^{1/2}(\theta_0) \right] = P \left[ A, \hat{\theta} \in B_r(\theta_0) \cap A, \hat{\theta} \notin B_r(\theta_0) \right]
\]

\[
\leq P \left[ A, \hat{\theta} \in B_r(\theta_0) \right] + P \left[ A, \hat{\theta} \notin B_r(\theta_0) \right]
\]

\[
\leq P \left[ \hat{\theta} \in B_r(\theta_0) \right] + P \left[ \hat{\theta} \in \Theta \setminus B_r(\theta_0) \right]
\]

\[
\leq P \left[ \hat{\theta} \in B_r(\theta_0) \right] + P \left[ \inf_{\theta \in \Theta \setminus B_r(\theta_0)} M_n^{1/2}(\theta) \leq M_n^{1/2}(\theta_0) \right]
\]

\[
= I + II
\]

If \( II \to 0 \), then we have for any \( r > 0 \) with probability tending to one, \( \hat{\theta} \in B_r(\theta_0) \) as \( n \to \infty \) which completes the proof. Therefore, we will show that \( II \to 0 \) as \( n \to \infty \).

Note that

\[
M(\theta) = E \left[ \{Y_i - g_0(X_i, \theta)\}^2 | U = u_0 \right]
\]

is the probability limit of (3). By adding and subtracting \( M(\theta) \) and \( M(\theta_0) \) into \( II \) we get

\[
II \leq P \left[ A1 + A2 \geq \inf_{\theta \in \Theta \setminus B_r(\theta_0)} M_n^{1/2}(\theta) - M_n^{1/2}(\theta_0) \right]
\]

where

\[
A1 = \sup_{\theta \in \Theta} |M_n^{1/2}(\theta) - M_n^{1/2}(\theta_0)|
\]

\[
A2 = |M_n^{1/2}(\theta_0) - M_n^{1/2}(\theta_0)|.
\]

Similar to (Ichimura (1993), pg89) by condition C3 we have

\[
\inf_{\theta \in \Theta \setminus B_r(\theta_0)} M_n^{1/2}(\theta) - M_n^{1/2}(\theta_0) > \epsilon
\]

Therefore the proof reduces to showing that \( P[A1 > \epsilon/2] \) and \( P[A2 > \epsilon/2] \) converging to zero as \( n \to \infty \). Note that the first convergence will imply the second. Consider \( A1 \). Let

\[
\tilde{M}_n(\theta) = (nh_2)^{-1} \sum_{i=1}^n \{Y_i - g_0(X_i, \theta)\}^2 K_2 \left( \frac{U_i - u_0}{h_2} \right)
\]
and $M^*(\theta) = E \left[ \tilde{M}_n(\theta) \right]$. Then we have

\[
A_1 \leq \sup_{\theta \in \Theta} |M_n^{1/2}(\theta) - \tilde{M}_n^{1/2}(\theta)| + \sup_{\theta \in \Theta} |\tilde{M}_n^{1/2}(\theta) - M^{1/2}(\theta)| + \sup_{\theta \in \Theta} |M^{1/2}(\theta) - M^*(\theta)| \\
= A_{11} + A_{12} + A_{13}.
\]

Consider $A_{11}$. Using the general result

\[
\left( \sum_{i=1}^{n} w_i a_i^2 \right)^{1/2} - \left( \sum_{i=1}^{n} w_i b_i^2 \right)^{1/2} \leq \sum_{i=1}^{n} w_i (a_i - b_i)^2
\]

and letting

\[
w_i = \frac{1}{nh_2} K_2 \left( \frac{U_i - u_0}{h_2} \right) \\
a_i = Y_i - \hat{g}_{u_0}(\theta^T X_i, \theta) \\
b_i = Y_i - g_0(X_i, \theta)
\]

we get

\[
|M_n^{1/2}(\theta) - \tilde{M}_n^{1/2}(\theta)|^2 \leq \sum_{i=1}^{n} \frac{1}{nh_2} K_2 \left( \frac{U_i - u_0}{h_2} \right) [g_0(X_i, \theta) - \hat{g}_{u_0}(\theta^T X_i, \theta)]^2
\]

\[
\leq \sup_{t,\theta} \left| g_0(X_i, \theta) - \hat{g}_{u_0}(\theta^T X_i, \theta) \right|^2 \frac{1}{nh_2} K_2 \left( \frac{U_i - u_0}{h_2} \right) .
\]

If $nh_2 \to \infty$ as $n \to \infty$ then by the consistency result of $\hat{g}_{u_0}(t, \theta)$ in Lemma 1 gives us

\[
\sup_{t,\theta} \left| M_n^{1/2}(\theta) - \tilde{M}_n^{1/2}(\theta) \right| = o_p(1).
\]

Now consider $A_{12}$. Note that

\[
|\tilde{M}_n^{1/2}(\theta) - M^{1/2}(\theta)|^2 \leq |\tilde{M}_n(\theta) - M^*(\theta)|.
\]

Let

\[
W_i(\theta) = \left( Y_i - g_0(X_i, \theta) \right)^2 K_2 \left( \frac{U_i - u_0}{h_2} \right) .
\]

Now we need to show that

\[
\sup_{\theta \in \Theta} \left| \frac{1}{nh_2} \sum_{i=1}^{n} W_i(\theta) - EW_i(\theta) \right| \overset{p}{\to} 0.
\]

This type of convergence results are established in Andrews (1987). By verifying assumption A1,B1,B2 and A4 of Andrews (1987) we see that $A_{12} \to 0$ in probability.

Finally, consider $A_{13}$. A simple calculation yields $M^*(\theta) = M(\theta) + O(h^2)$. By our assumptions in C9 we see that the order term is uniformly bounded in $\theta$ which shows that $A_{13} \to 0$ in probability. This completes the proof.
References


