Sieve Empirical Likelihood Ratio Tests
for Nonparametric Functions*

By Jianqing Fan and Jian Zhang

December 26, 2000

Abstract

Generalized likelihood ratio statistics have been proposed in Fan, Zhang and Zhang (2001) as a generally applicable method for testing nonparametric hypotheses concerning about nonparametric functions. The likelihood ratio statistics are constructed based on the assumption that the distributions of stochastic errors are in a certain parametric family. We extend their work to the case where the error distribution is completely unspecified via newly proposed sieve empirical likelihood ratio tests. The approach is also applied to test conditional estimating equations on the distributions of stochastic errors. It is shown that the proposed sieve empirical likelihood ratio statistics follow asymptotically rescaled $\chi^2$-distributions, with the scale constants and the degrees of freedom being independent of the nuisance parameters. This demonstrates that the Wilks phenomenon observed in Fan, Zhang and Zhang (2001) continues to hold under more relaxed models and a larger class of techniques. The asymptotic power of the proposed test is also derived, which achieves the optimal rate for nonparametric hypothesis testing. The proposed approach has two advantages over the generalized likelihood ratio method: it requires only to specify some conditional estimating equations rather than the entire distribution of the stochastic error and the procedure adapts automatically to unknown error distributions including heteroscedasticity.

Key words: Nonparametric test, empirical likelihood, Wilks' theorem, varying-coefficient models.


---

Fan's research was partially supported by NSF grant DMS-0196041 and the RGC grant CUHK-4299/00P of HKSAR. J. Zhang's research was partially supported by the National Natural Science Foundation of China and a grant from the research programme in EURANDOM, Netherlands. The authors would like to thank Ms. Anna Liu for her help in preparing this paper.
1 Introduction

Over the last two decades, many computationally intensive nonparametric techniques and theory have been flourishedly developed to exploit possible hidden structures and to reduce modeling biases of traditional parametric methods. Methods such as local polynomial fitting, spline approximations and orthogonal series expansions as well as dimensionality reduction techniques have been studied in great depth in various statistical contexts. Yet, there are no generally applicable methods available for the inferences in nonparametric models. Various efforts have been made in the literature on nonparametric hypothesis testing. See for example Bickel and Ritov (1992), Eubank and Hart (1992), Härdle and Mammen (1993), Azzalini and Bowman (1993), Fan (1996), Spokoiny (1996), Ingliot and Ledwina (1996), Kallenberg and Ledwina (1997), among others. For an overview, see the recent book by Hart (1997). However, most of these methods focus only on the one-dimensional nonparametric regression problem. They are difficult to be extended to multivariate semiparametric and nonparametric models.

In an effort to derive a generally applicable testing procedure, for multivariate semiparametric and nonparametric models, Fan, Zhang and Zhang (2001) proposed generalized likelihood ratio tests. The work is motivated by the fact that the nonparametric maximum likelihood ratio test may not exist in many nonparametric problems. Further, even if it exists, it is not optimal even in the simplest nonparametric regression setting. Generalized likelihood ratio statistics, obtained by replacing unknown functions by reasonable nonparametric estimators, rather than the MLE as in the parametric setting, have several nice properties. In the varying-coefficient model

\[ Y = a_1(U)X_1 + \cdots + a_p(U)X_p + \varepsilon, \]  \hspace{1cm} (1.1)

where \((U, X_1, \cdots, X_p)\) are independent variables and \(Y\) is the response variable, Fan, Zhang and Zhang (2001) unveil the following Wilks phenomenon: The asymptotic null distributions are independent of nuisance functions and follow a \(\chi^2\)-distribution (in a generalized sense) for testing the homogeneity

\[ H_0 : a_1(\cdot) = \theta_1, \cdots, a_p(\cdot) = \theta_p \] \hspace{1cm} (1.2)

and for testing the significance of variables such as

\[ H_0 : a_1(\cdot) = a_2(\cdot) = 0. \] \hspace{1cm} (1.3)

In other words, the generalized likelihood ratio statistic \(\lambda_n\) follows asymptotically a rescaled \(\chi^2\)-distribution in the sense that \((2b_n)^{-1/2}(r_K \lambda_n - b_n) \xrightarrow{D} N(0,1)\) for a sequence \(b_n \to \infty\) and a constant \(r_K\). We will use the notation \(r_K \lambda_n \overset{D}{\sim} \chi^2{b_n}\) to denote the result. The significance of the result is that the scale constant \(r_K\) and the degree of freedom \(b_n\) are independent of nuisance parameters, such as the joint density of \((U, X_1, \cdots, X_p)\) and the parameters \(\theta_1, \cdots, \theta_p\) in (1.2) and the functions \(a_1(\cdot), \cdots, a_p(\cdot)\) in (1.3). This Wilks phenomenon is the key to the success of the classical maximum likelihood ratio tests for parametric problems. With the above newly discovered Wilks phenomenon in nonparametric models, the P-values can easily be computed by using either the asymptotic distributions or simulations via fixing nuisance parameters or functions under the null hypothesis at certain values of interest. Further, Fan, Zhang and Zhang (2001) show that the resulting tests are asymptotically optimal in the sense of Ingster (1993).

The idea of the above generalized likelihood method is widely applicable in semiparametric and nonparametric models. It is easy to use because of the Wilks phenomenon and is powerful as it achieves the optimal rate of convergence. Yet, one needs to specify the parametric form of the error distribution such
as \( \varepsilon \) in (1.1) in order to construct the generalized likelihood ratio statistic. While the procedure based on the normal likelihood may be still applicable to the case where the distribution of \( \varepsilon \) is homoscedastic, it may not be efficient. When the error distribution is heteroscedastic with the variance \( \text{var}(\varepsilon|U) = \sigma^2(U) \), the construction of the generalized likelihood ratio test statistic needs the knowledge of the variance function \( \sigma^2(\cdot) \). This motivates us to propose the sieve empirical likelihood ratio test statistic for handling the case where the exact form of the error distribution is unknown, but some qualitative traits of the distribution is known. A popular model is to assume

\[
E[G(\varepsilon)|U] = 0
\]  

where \( G = (G_1, \cdots, G_{k_o})' \) is a \( k_o \)-dimensional function (see Owen, 1988; Newey, 1993; Zhang and Gijbels, 1999). This is a much less restrictive assumption than a parametric form on the distribution of \( \varepsilon \). In particular, when the conditional distribution of \( \varepsilon \) given \( U \) is symmetric about 0, we may choose a sequence of \( k_o \) grid points, say, \( 0 = s_0 < s_1 < \cdots < s_{k_o} \) and take

\[
G_k(\varepsilon) = I(\varepsilon \in [s_{k-1}, s_k]) - I(-\varepsilon \in [s_{k-1}, s_k]), \quad 1 \leq k \leq k_o
\]  

or a smoother version of the function \( G_k \). Note that as \( \max_{1 \leq k \leq k_o}(s_k - s_{k-1}) \to 0 \), \( k_o \to \infty \), these restrictions are essentially the same as the symmetric assumption.

A few questions related to the sieve empirical likelihood ratio test arise naturally. First of all, it is not clear how to construct an empirical likelihood in the nonparametric setting. Secondly, it is not obvious whether a particular construction of the empirical likelihood ratio statistic will follow the Wilks' type of result. Thirdly, it is not granted that the resulting test statistic is asymptotically optimal in the sense of Ingster (1993). Finally, it remains unknown whether the empirical likelihood ratio statistics will adapt to unknown distribution of \( \varepsilon \) including heteroscedasticity. These issues are poorly understood and the new phenomena need to be discovered.

The technical derivations for sieve empirical likelihood ratio tests are very involved. To ease some of the technical burden, we choose the varying coefficient model (1.1) for our investigation. The model arises from various contexts and has been widely used. For example, in many biomedical studies, it is frequently encountered the issue such as the extent to which the effect of exposure variables on the response variable changes with the level of a confounding covariate (e.g. age). See, for example, Cleveland, Grosse and Shyn (1991), Hastie and Tibshirani (1993) and Carroll, Ruppert, and Welsh (1998). In longitudinal studies, investigators often want to exam how the effects of covariates on response variables change over time (Brumback and Rice, 1998, Wu, Chiang and Hoover 1998). In nonlinear time series, the model allows different autoregressive model for different regimes of state variables (Chen and Tsay, 1993; Cai, Fan and Yao, 2000). It includes the thresholded autoregressive model (Tong, 1990) as a specific example. The model has successfully been applied by Hong and Lee (1999) to the inference and forecast of exchange rates. Thus, our study in model (1.1) has direct implications on the above problems.

For the varying coefficient model (1.1), it arises frequently whether the coefficient functions are really varying or whether certain covariates are statistically significant. This leads to the problem of testing for homogeneity (1.2) or the problem of testing for significance such as the problem (1.3). As to be explained at the end of section 2, these problems can be reduced to the problem of testing against a specific null hypothesis:

\[
H_0: a_1(\cdot) = a_{10}(\cdot), \cdots, a_p(\cdot) = a_{p0}(\cdot),
\]
for some given functions $a_{10}, \ldots, a_{p0}$. Our approach is to first construct the local linear estimator of the coefficient functions $a_1, \ldots, a_p$ via a local version of the empirical likelihood, then substitute the estimate into a special sieve empirical likelihood (see Zhang and Gijbels, 1999; Liu and Zhang, 2000). This allows us to form the empirical likelihood ratio statistics. We will show that the proposed sieve empirical likelihood procedures follow the Wilks type of results under more relaxed assumptions on the error distribution of $\varepsilon$. This provides a useful extension of the results given by Fan, Zhang and Zhang (2001). Note that our procedure is very different from that of Kitamura (1997) in which he considered testing problems for finite dimensional parameters in weakly dependent processes. He first used the local (blocking) approximation to construct a global estimating equation, then apply Owen’s procedure directly.

Our empirical likelihood ratio method applies also to the nonparametric tests on density functions. As an illustration without introducing new statistical setting, we regard the constraints (1.4) as a null hypothesis. We will demonstrate that the Wilks type of phenomenon continues to hold for this nonparametric testing problem.

When $p = 1$ and $X \equiv 1$ and the coefficient function $a_1(\cdot) \equiv \theta$, under the constraints (1.4) and (1.5), the model becomes a one-sample symmetric location model which is well studied, for instance, by Hettmansperger (1984) and Bickel et al. (1993). In Section 2, we find that for this special case, the first step in our procedure essentially makes the information on the stochastic error to be efficiently used (see, Owen, 1988; Liu and Zhang, 2000). Moreover, the second step makes the likelihood ratio statistic adaptive to heteroscedasticity. As a result, our procedure has two advantages over the parametric model on the error distribution. Firstly, it requires only some conditional estimating equations such as (1.4) rather than the whole distribution of the stochastic error. Secondly, the asymptotic null distribution of the sieve empirical likelihood ratio statistic asymptotically follows a rescaled $\chi^2$-distribution. The scaling constant and the degree of freedom are independent of the conditional distribution of $\varepsilon$ even if the stochastic error is heteroscedastic in $U$. The procedure and results can be easily generalized to a more general constrained regression model in Zhang and Gijbels (1999).

The paper is organized as follows. In Section 2, the sieve empirical likelihood ratio statistics are introduced for testing the goodness-of-fit of the estimating equations and for testing some simple and composite null hypotheses. In Section 3, the asymptotic null and non-null distributions of these statistics are derived. The technical conditions and the proofs are deferred to Section 4.

## 2 Sieve Empirical likelihood

It is more convenience to work on the matrix notation for model (1.1):

$$Y = A^r(U)X + \varepsilon,$$

where $Y$ is the response, $U \in \Omega \subset R^d$ (with $\Omega$ bounded) and $X \in R^p$ are covariates, $\varepsilon$ is the stochastic error and $A(U) = (a_1(u), \ldots, a_p(u))$ is the vector of varying coefficients. Let $\{(Y_i, X_i, U_i)\}_{i=1}^n$ be an iid random sample from the model (2.1) with the restriction (1.4). According to Owen’s procedure (Owen, 1988), to construct an empirical likelihood which can identify an infinite dimensional parameter such as $A(u)$ in (2.1), we need to establish an infinite number of unconditional estimating equations. Such a likelihood is often theoretically intractable. To overcome this difficulty, Zhang and Gijbels (1999) proposed a general procedure to build a sieve empirical likelihood via the local approximation. For the model (2.1) the procedure...
consists of two steps: First, for each \( U_j \) construct \( n \) local empirical likelihoods which can locally identify \( A(u) \), \( u \approx U_j \). These local empirical likelihoods lead to a weighted approximation of the logarithm of the conditional probability mass \( dP(Y, X|U = U_j, Y_j, X_j) \). Then a log-likelihood is obtained simply by summing up all of these approximated logarithms. In the first step, we will use the local linear approximation of the nonparametric coefficient functions \( A(\cdot) \) (see Fan and Zhang, 1999; Cai, Fan and Li, 2000). In other words, in a neighborhood around a given point \( u_0 \), approximate \( A(\cdot) \) by

\[
A(u) \approx A(u_0) + A'(u_0)(u - u_0), \quad \text{for } u \approx u_0.
\]

Thus, around the point \( u_0 \), the model (2.1) and the restriction (1.4) can be written respectively as

\[
Y \approx \beta_A(u_0)^T Z(X, (U - u_0)/h) + \varepsilon, \quad \text{for } U \approx u_0,
\]

\[
E[G(Y - \beta_A(u_0)^T Z(X, (U - u_0)/h)|U = u] \approx 0, \quad \text{for } u \approx u_0
\]

(2.2)

where \( \beta_A(u_0) = (A^T(u_0), hA^T(u_0))^T \) and \( Z(X, t) = (X^T, tX^T)^T \). This is indeed a local linear model. To incorporate the local linear model, let \( h \) represent the size of the local neighborhood where the approximation is valid and \( K \) be a weight function, which is a symmetric probability density function. Let \( p_i, i = 1, \cdots, n \) be the conditional empirical probability mass of \( (X, Y) \) given \( U = u_0 \), putting on the \( i \)-th data point \( (X_i, Y_i) | (i = 1, \cdots, n) \). Suppose that given \( U, \varepsilon \) and \( X \) are independent. Then, the conditional constraints (2.2) can be translated into the following unconditional estimating equation

\[
\sum_{i=1}^n p_i G_{ih}(u_0, \beta_A(u_0)) = 0,
\]

where

\[
G_{ih} = G_{ih}(u_0, \beta) = G(Y_i - \beta^T Z(X_i, (U_i - u_0)/h)) \otimes Z(X_i, (U_i - u_0)/h)
\]

with \( \otimes \) being the Kronecker product, \( \beta = (A^T, hB^T)^T, A = (a_1, \cdots, a_p)^T \) and \( B = (b_1, \cdots, b_p)^T \). To see why we need an extra factor \( Z(X_i, (U_i - u_0)/h) \) in the unconditional estimating function \( G_{ih} \), we let \( G(\varepsilon) = \varepsilon \) temporarily. It is actually based on a well-known fact that in the linear model the product of the residual and the covariates is a good estimating equation for the parameter \( \beta_A \). This leads to the following estimating equation

\[
\sum_{i=1}^n p_i(Y_i - \beta^T Z(X_i, (U_i - u_0)/h))Z(X_i, (U_i - u_0)/h) = 0.
\]

In light of this fact, for a general \( G \), we should build the estimating equation by multiplying each components of \( G \) by the covariate vector \( Z(X_i, (U_i - u_0)/h) \), which admits the form \( G_{ih} \).

Thus, following Owen (1988), the local empirical log-likelihood function of \( \beta \) is defined by

\[
I(\beta, u_0) = \sup \{ \sum_{i=1}^n w_h(U_i, u_0) \log p_i : p_i \geq 0, 1 \leq i \leq n, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i G_{ih}(u_0, \beta) = 0 \}
\]

where \( w_h(U_i, u_0) = K_h(U_i - u_0)/\sum_{m=1}^n K_h(U_m - u_0) \) with \( K_h(\cdot) = K(\cdot/h)/h \). By the Lagrange multiplier method, we obtain

\[
I(\beta, u_0) = \sum_{i=1}^n w_h(U_i, u_0) \log w_h(U_i, u_0) - \sum_{i=1}^n w_h(U_i, u_0) \log(1 + \alpha_h(U_i, u_0) G_{ih}),
\]

5
where \( \alpha_n(u_0, \beta) \) satisfies
\[
\sum_{i=1}^{n} w_h(U_i, u_0) \frac{G_{th}(u_0, \beta)}{1 + \alpha_n(u_0, \beta)G_{th}(u_0, \beta)} = 0. \tag{2.3}
\]

Define the estimate of \( \beta \) by
\[
\hat{\beta}(u_0) = \text{argmax}_\beta l(\beta, u_0). \tag{2.4}
\]

The first \( p \) components, denoted by \( \hat{A}(u_0) \), give an estimate of \( A(u_0) \), and the remaining components estimate \( hA'(u_0) \). An approximate empirical likelihood, called the sieve empirical likelihood for the nonparametric function \( A \) can be introduced by adding the logarithm of the conditional likelihood at each data point:
\[
l(A|G) = \sum_{j=1}^{n} l(\beta_A, U_j).
\]

The name “sieve” originates from the following two facts: (1) in the above procedure \( E[G(\varepsilon)|U = U_j] \) is a sieve approximation to the constraints (1.4); (2) \( l(\beta_A, U_j) \) is the weighted approximation of the logarithm of the conditional probability mass \( dP_{(Y, X)|U = U_j}(Y, X) \). See Zhang and Gijbels (1999) for a more detailed explanation. Motivated by Fan, Zhang and Zhang (2001), we define the logarithm of the sieve empirical likelihood under the nonparametric model (2.1) with constraints (1.4) by substituting \( \beta = \hat{\beta} \) into \( l(A|G) \), leading to
\[
l(\theta|G) = \sum_{j=1}^{n} l(\hat{\beta}(U_j), U_j). \]

We now consider the nonparametric test concerning about the density function of \( \varepsilon \). As a specific application of the sieve empirical likelihood, we consider the testing
\[
H_{0G} : E[G(\varepsilon)|U] = 0, \tag{2.5}
\]
where \( G \) is given in (1.4). Without the constraint (1.4), following the above derivations, the corresponding logarithm of the sieve empirical likelihood is
\[
l(\theta|N) = \sum_{j=1}^{n} \sum_{i=1}^{n} w_h(U_i, U_j) \log w_h(U_i, U_j).
\]

Thus, we can construct a goodness-of-fit test of the hypothesis (2.5) based on the following logarithm of the sieve empirical likelihood ratio:
\[
l(G) = l(\theta|G) - l(\theta|N),
\]
\[
= \sum_{j=1}^{n} \sum_{i=1}^{n} w_h(U_i, U_j) \log (1 + \hat{\alpha}(U_j)^* G_{th}(U_j, \hat{\beta})) \tag{2.6}
\]

where \( \hat{\alpha}(u) = \alpha_n(u, \hat{\beta}) \).

Next, we consider the sieve likelihood ratio test for the nonparametric coefficient function \( A(\cdot) \) under the restriction (1.4). In the varying coefficient model (2.1), we ask naturally whether the coefficient is really varying or whether certain covariates are statistically significant. This leads to the parametric null hypothesis:
\[
H_{0A} : A(\cdot) = \theta. \]
More generally, we wish to test the composite null hypothesis, which involves nuisance functions $A_2(\cdot)$:

$$H_{0u} : A_1 = A_{10} \quad \iff \quad H_{1u} : A_1 \neq A_{10} \quad (2.7)$$

with $A_2(\cdot)$ completely unknown. This problem includes the test of significance (1.3) under model (1.1) as a specific example. Here we write

$$A_0(u_0) = \begin{pmatrix} A_{10}(u_0) \\ A_{20}(u_0) \end{pmatrix}, \quad \text{and} \quad A(u_0) = \begin{pmatrix} A_1(u_0) \\ A_2(u_0) \end{pmatrix},$$

with $A_{10}(u)$ and $A_1(u)$ being $p_1 < p$ dimensional. To construct the likelihood ratio statistic for $H_{0u}$, we introduce the following notation:

$$\beta_{2A}(u_0) = (A_2^T(u_0), hA_2^{tr}(u_0)), \quad \beta_2 = (A_2^T, hB_2^T)^T, \quad \beta^* = (A_{10}^T(u_0), A_2^T, hA_{10}^{tr}(u_0), hB_2^T)^T.$$ 

Let

$$\hat{\beta}_2(u_0) = (\hat{A}_2^T, h\hat{B}_2^T)^T = \arg\max_{\beta_2} l(\beta^*, u_0),$$

$$\hat{\beta}^*(u_0) = (\hat{A}_{10}^T(u_0), \hat{A}_2^T, h\hat{A}_{10}^{tr}(u_0), h\hat{B}_2^T)^T$$

and the corresponding $\hat{\alpha}^*(u_0)$ be defined by

$$0 = \frac{1}{n} \sum_{i=1}^n w_i(U_i, u_0) \frac{G_{ih}(u_0, \hat{\beta}^*(u_0))}{1 + \hat{\alpha}^{tr}(u_0)G_{ih}(u_0, \hat{\beta}^*(u_0))}.$$ 

Then, the sieve empirical likelihood ratio statistic for $H_{0u}$ can be written as

$$l(H_{0u}|G) = l(\Theta_{02}|G) - l(\Theta|G). \quad (2.8)$$

with

$$l(\Theta_{02}|G) = \sum_{j=1}^n l(\hat{\beta}^*(U_j), U_j).$$

The sieve empirical likelihood ratio test for the semiparametric model that $A(\cdot)$ has a certain parametric form such as the linear model can be constructed analogously. Similarly to Fan, Zhang and Zhang (2001), the asymptotic null distributions of the sieve likelihood ratio statistics for composite null hypotheses can be derived from those for simple hypotheses (see the next paragraph). This motivates us to consider

$$H_{0s} : A = A_0 \quad \iff \quad H_{1s} : A \neq A_0 \quad (2.9)$$

for a given $A_0$. Analogously to $l(H_{0u})$, we can construct the following likelihood ratio statistic:

$$l(H_{0s}|G) = l(A_0|G) - l(\Theta|G)$$

$$= \sum_{j=1}^n \sum_{i=1}^n w_i(U_i, U_j) \log(1 + \alpha_n(U_j, \beta_0)^*G_{ih}(U_j, \beta_0)) - l(\Theta|G) \quad (2.10)$$

where $\beta_0$ denotes $\beta_{A_0}$. Note that when $A_0$ in $H_{0s}$ is known, we can assume, without loss of generality, that $A_0 \equiv 0$. This can be accomplished through a simple transformation $A^* = A - A_0$. With this transformation, (2.9) is equivalent to

$$H'_{0s} : A^* \equiv 0 \quad \iff \quad H'_{1s} : A^* \neq 0. \quad (2.11)$$
We opt for general $A_0$, since the results have implications on the composite null hypotheses. To appreciate this, consider the composite null hypothesis testing problem:

$$H_0 : A \in A_0 \iff A \notin A_0,$$  \hspace{1cm} (2.12)

where $A_0$ is a set of functions. Let $l(A_0|G)$ be the sieve empirical likelihood under the hypothesis $H_0$ in (2.12). Then, the sieve empirical likelihood ratio statistic is simply

$$\lambda_n = l(A_0|G) - l(\Theta|G).$$

Let $A'_0$ denote the true value of the parameter function $A$. Consider the fabricated testing problems with the simple null hypotheses:

$$H'_0 : A = A'_0, \quad \iff \quad H'_1 : A \neq A'_0 \hspace{1cm} (2.13)$$

and

$$H'_0 : A = A'_0, \quad \iff \quad H'_1 : A \in A_0. \hspace{1cm} (2.14)$$

Let $l(A'_0|G)$ be the sieve empirical likelihood under $H'_0$. Then, the sieve empirical likelihood ratio statistic for (2.12) can be written as

$$\lambda_n = \lambda(A'_0|G) - \lambda^*(A'_0|G),$$

where $\lambda(A'_0|G) = l(A'_0|G) - l(\Theta|G)$ is the sieve empirical likelihood ratio statistic for the problem (2.13) and $\lambda^*(A'_0|G) = l(A'_0|G) - l(A_0|G)$ is the sieve empirical ratio test for the problem (2.14). Thus, the asymptotic representation of $\lambda_n$ follows directly from those of $\lambda(A'_0)$ and $\lambda^*(A'_0)$, which admits the form given by (2.10).

3 Asymptotic theory

3.1 Asymptotic expansions

In order to obtain the properties of the sieve empirical likelihood ratio statistics in (2.6) and (2.10), we first develop some uniform asymptotic representations for the local sieve empirical likelihood estimator $\hat{B}$ and the Lagrange multiplier $\hat{\alpha}$ in (2.3) and (2.4). These results are the generalizations of Liu and Zhang (2000). They also indicate the performance of the sieve empirical likelihood estimator. Using these results we will establish the asymptotic representations for $l(G)$ and $l(H_{0U}|G)$ in (2.6) and (2.10). For the simplicity of presentation, we assume $G$ is differentiable. Let $f(u_0)$ be the density of $U$ at the point $u_0$. Set

$$D(u_0) = -E[\frac{\partial G(\varepsilon)}{\partial \varepsilon}[U = u_0], \quad V(u_0) = E[G(\varepsilon)G^*(\varepsilon)[U = u_0],$$

$$\Gamma(u_0) = E[XX'U = u_0]f(u_0), \quad S = \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}, \quad \mu_2 = \int t^2 K(t)dt.\$$

$$\eta_i(u_0) = -(D(u_0)^*V(u_0)^{-1}D(u_0))^{-1}D^*(u_0)V^{-1}(u_0)G(\varepsilon_i),$$

$$C(u_0) = V^{-1}(u_0) - V^{-1}(u_0)(D^*(u_0)V^{-1}(u_0)D(u_0))^{-1}D(u_0)D^*(u_0)V^{-1}(u_0),$$

$$\varepsilon_i = \gamma_i - A^*(U_i)X_i.$$  

Theorem 1 Suppose that the conditions (K0), (U0), (A1)-(A10) and (B1)-(B5) in Section 4.1 hold and that the underlying $A(u)$ is linear or has the twice continuous derivative and satisfies the condition (B6). If
there exist some positive constants $b_0, b_1$ and $\eta < 1/2$ such that $b_0 \leq h^n \eta \leq b_1$, then uniformly for $u_0 \in \Omega$,

\[
\hat{\beta}(u_0) = \beta(u_0) + \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0) \left( \frac{\Gamma^{-1}(u_0) X_i}{\mu_2^{-1} \Gamma^{-1}(u_0) X_i (U_i - u_0)/h} \right) \eta(u_0)(1 + o_p(h^{1/2})) + O_p(h^2),
\]

\[
\hat{\alpha}(u_0) = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0) \{ C(u_0) G(\varepsilon_i) \} \otimes \left( \frac{\Gamma^{-1}(u_0) X_i}{\mu_2^{-1} \Gamma^{-1}(u_0) X_i (U_i - u_0)/h} \right)(1 + o_p(h^{1/2})) + O_p(h^2).
\]

As a consequence of Theorem 1, we have the following asymptotic uniform expansion:

\[
\hat{A}(u_0) - A(u_0) = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0) \Gamma^{-1}(u_0) X_i \eta(u_0)(1 + o_p(h^{1/2})) + O_p(h^2).
\]

The asymptotic normality of the sieve empirical likelihood estimator follows easily from the above asymptotic expansion.

In Theorem 1, the requirement that $G$ is differentiable can be relaxed by imposing some entropy conditions on $G$ and by assuming $E[G(e - t)|U = u_0]$ is twice continuously differentiable in $t$. In this case $D(u_0)$ should be replaced by $-\{\partial E[G(e - t)|U = u_0]/\partial t\}|_{t=0}$. Similar to Liu and Zhang (2000), we can show that the asymptotic efficiency of $\hat{A}(u_0)$ is increasing in $D(u_0)^* V(u_0)^{-1} D(u_0)$. In particular, in the setting of symmetric location model mentioned in Section 1, we can find a sequence of $G$, say $\{G^{(k)}\}$ such that the corresponding $\hat{A}(u_0)$ is asymptotically adaptive to the unknown conditional density of $e$ given $U = u_0$. In practice, to save the computational effort, we prefer to choose a $G$ with a small $k_0$ and a relatively larger $D(u_0)V(u_0)^{-1}D(u_0)$.

We now give the asymptotic representations for the sieve empirical likelihood ratio statistics $l(G)$ and $l(H_{0G}|G)$. The results indicate that they admit a generalized quadratic form. To facilitate the notation, the following notation is introduced. Let

\[
\phi_{ikh}(U) = K_h(U_i - U)K_h(U_k - U)C(U)(1 + (U_i - U)(U_k - U)\mu_2^{-1} h^{-2})X_k^{-1}(U)X_k f^{-1}(U),
\]

\[
K^*(s) = \int K(t)K(s + t)(1 + t(s + t)\mu_2^{-1})dt,
\]

\[
\Phi_{ikh} = E[\phi_{ikh}(U)|U_i, U_k, X_i, X_k],
\]

\[
= K_i(\mu_{ikh}(U_k - U_i)C(U_i)X_i^{-1}(U_i)X_k(1 + O_p(h))),
\]

\[
T_n = \frac{1}{n(n-1)} \sum_{i \neq k} G^*(\varepsilon_i) \Phi_{ikh} G(\varepsilon_k).
\]

Similarly, we define

\[
q_{ikh}(U) = K_h(U_i - U)K_h(U_k - U) V^{-1}(U)X_i^{-1}(U)X_k(1 + (U_i - U)(U_k - U)\mu_2^{-1} h^{-2}) f^{-1}(U),
\]

\[
Q_{ikh} = E[q_{ikh}(U)|U_i, U_k, X_i, X_k],
\]

\[
T_n^* = \frac{1}{n(n-1)} \sum_{i \neq k} G^*(\varepsilon_i) (Q_{ikh} - \Phi_{ikh}) G(\varepsilon_k).
\]

Then, we have the following result.

**Theorem 2** Suppose the conditions of Theorem 1 hold. Then under $H_{0G}$,

\[
l(G) = (k_n - 1)/h \int K^2(t)(1 + t^2\mu_2^{-1})dt + (1 + o_p(h^{1/2}))nT_n + o_p(h^{-1/2});
\]

\[
l(H_{0G}|G) = (k_n - 1)/h \int K^2(t)(1 + t^2\mu_2^{-1})dt + (1 + o_p(h^{1/2}))nT_n^* + o_p(h^{-1/2});
\]

\[9\]
and under $H_{04}$, if $A_0$ is linear or $nh^{3/2} \to 0$, then
\begin{equation}
2l(H_{04} \mid G) = \frac{p|\Omega|}{h} \int K^2(t)(1 + t^2 \mu_2^{-1})dt + (1 + o_p(h^{1/2}))nT_n^* + o_p(h^{-1/2}),
\end{equation}
(3.3)
where $|\Omega|$ is the length of the support $\Omega$ of the density $f$.

Note that if there are no components in $A$, then under $H_{04}$ the factor $k_0 - 1$ in (3.2) should be $k_0$, since we cost $p$ degrees of freedom to estimate them when there are $p$ components in $A$.

3.2 Asymptotic null distribution

With the asymptotic representations, we are now ready to derive the asymptotic distributions of the test statistics $l_G$ and $l(H_{04} \mid G)$. Like in the parametric case for the stochastic error $\varepsilon$ (see Fan, Zhang and Zhang, 2001), under the null hypotheses the sieve empirical likelihood ratio statistics in (2.6), (2.8) and (2.10) are asymptotically $\chi^2$-distributed and their degrees of freedom are independent of the nuisance parameters such as $A$, $G$, and the distribution of $\varepsilon$.

**Theorem 3** Under $H_{04}$ and the conditions of Theorem 1, for $k_0 > 1$, we have $\tau_K l_G \overset{a}{\sim} \chi^2_n$ with
\[ \tau_K = \frac{2K^*(0)}{\int K^*(s)^2 ds}, \quad b_n = \frac{(k_0 - 1)p|\Omega|c_K}{h}, \]
where $K^*(s)$ is defined in (3.1), $c_K = K^*(0)^2 / \int K^*(s)^2 ds$. For $k_0 = 1$, we have $\tau_K l_G = o_p(1)$.

**Remark 3.1** If $K(t)$ has support $[-1, 1]$, and if $K(t)$ and $|t|K(t)$ are concave on $t \in [-1, 1]$, then by the same argument used in the Sherman inequality (see, Farrell, 1985, pp.343), we have
\[ |K^*(s)| \leq \int K(t)K(s + t)dt + \mu_2^{-1} \int |t|K(t)|s + t|K(s + t)dt \leq K^*(0). \]
Thus when $K^*(s) \geq 0, s \in [-1, 1], \tau_K \geq 2$. In particular, when $K$ is the uniform kernel function, $\tau_K = 2.8176$ and $c_K = 1.0566$; when $K$ is the Epanechnikov kernel function, $\tau_K = 2.5154$ and $c_K = 1.2936$.

The next theorem presents the asymptotic null distribution of $l(H_{04} \mid G)$.

**Theorem 4** Suppose that the conditions of Theorem 1 hold. Then under $H_{04}$, $\tau_K l(H_{04} \mid G) \overset{a}{\sim} \chi^2_n$; and under $H_{04}$, if $nh^{3/2} \to 0$, then $\tau_K l(H_{04} \mid G) \overset{a}{\sim} \chi^2_{n+2}$ where $b^*_n = p|\Omega|c_K/h$ and $b^*_n = p|\Omega|c_K/h$ with $c_K$ and $\tau_K$ defined in Theorem 3 and $p_1$ being the dimensionality of $A_{10}$ in (2.7).

Theorems 3 and 4 indicate that the sieve empirical likelihood ratio statistics continue to apply to the case where the distribution of the stochastic error $\varepsilon$ is completely unknown and furthermore there are many nuisance parameters in null hypotheses. In particular, the stochastic errors are allowed to be heteroscedastic and unknown. This is a useful generalization of the results in Fan, Zhang and Zhang (2001) where the distribution of $\varepsilon$ is essentially known. In particular, if the variance is heteroscedastic with $\text{var}(\varepsilon \mid U) = \sigma^2(U)$, they have to rely on the knowledge of $\sigma^2(\cdot)$ to construct the likelihood ratio statistics. This drawback is repaired by the empirical likelihood ratio method, while their Wilks phenomenon is inherited.
3.3 Asymptotic power

To demonstrate the effectiveness of the sieve empirical likelihood method, we consider, for simplicity, the test statistic for the problem (2.11) under the contiguous alternative \( A_n(\cdot) \to 0 \), with \( A_n''(\cdot) \) being bounded. That is, we allow the coefficient functions close to the null hypothesis, but is still in the class of functions with the bounded and continuous second derivatives. This is a much less restriction than the contiguous alternatives of form \( A_n(u) = a_n B_0(u) \) for a sequence \( a_n \to 0 \) and a given \( B_0 \), considered by many authors (e.g. Eubank and Hart, 1992, Eubank and LaRiccia, 1992, Hart, 1997, Inglot and Ledwina 1996). The latter implicitly assumes that \( A_n'(u) \to 0 \) and \( A_n''(u) \to 0 \), which are too restrictive for nonparametric applications.

We begin with the following notation. Let

\[
W_{1n}^* = \frac{1}{n} \sum_{i \neq k} K_n^*(U_i - U_k) G(\varepsilon_i) V^{-1}(U_k) X_i \Gamma^{-1}(U_k) X_k A(U_k)^T X_k \frac{\partial G(\varepsilon_k)}{\partial \varepsilon},
\]

(3.4)

\[
\Xi_i = \frac{\partial G(\varepsilon_i)}{\partial \varepsilon} - E[\frac{\partial G(\varepsilon_i)}{\partial \varepsilon} | U_i],
\]

\[
W_{2n}^* = \frac{1}{n} \sum_{i \neq k} K_n^*(U_i - U_k) \Xi_i^T V^{-1}(U_i) \Xi_i A(U_i)^T X_i X_i \Gamma^{-1}(U_k) X_k A(U_k),
\]

(3.5)

\[
W_{3n}^* = \frac{1}{n} \sum_{i \neq k} K_n^*(U_i - U_k) \Xi_i^T V^{-1}(U_k) E[\frac{\partial G(\varepsilon_k)}{\partial \varepsilon} | U_k] A(U_i) X_i X_i \Gamma^{-1}(U_k) X_k A(U_k).
\]

(3.6)

Then, following the same arguments used in Fan, Zhang and Zhang (2001), we can derive the asymptotic power \( l(H_{0.1}|G) \) through the next theorem.

**Theorem 5** Assume that \( A_0 \equiv 0 \) and that the underlying coefficient \( A = A_n \) has the twice continuous derivatives and satisfies \( n^2 E(A(U)^T X X^T A(U)) = O(1) \), \( \max_u \|A(u)\| \to 0 \) and \( \max_u \|A''(u)\| = O(1) \) as \( n \to \infty \). Assume that \( G \) is twice continuously differentiable. Then under the conditions of Theorem 1,

\[
2l(H_{0.1}|G) = \frac{E[|G|]}{\hat{h}} K^*(0) + nE(D(U)^T V^{-1}(U) D(U) A(U)^T X X^T A(U)) \{1 + o(1)\}
\]

\[
- \frac{nh^4}{4} E\{D(U)^T C(U) D(U) A''(U)^T X X^T A''(U)\}
\]

\[
\times \int \int t^2(s + t)^2 K(t) K(s + t) (1 + \mu_2^{-1} t(s + t)) dt ds (1 + o(1))
\]

\[
+(1 + o_p(h^{1/2})) \{T_{1n}^* + 2W_{1n}^* + W_{2n}^* + 2W_{3n}^* \} + o_p(h^{-1/2})
\]

where \( D, V, C \) and \( K^* \) are defined in Subsection 3.1.

Using the above result, similar to that in Fan, Zhang and Zhang (2001), it can easily be shown that the sieve empirical likelihood ratio can detect alternative with rate \( n^{-4/9} \) when \( h = c_n n^{-2/9} \) for some constant \( c_n \). This rate is optimal in the ordinary nonparametric regression setting.
4 Technical conditions and proofs

4.1 Technical conditions

Define

\[ A_n(u_0, \beta) = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0) G_{ih}(u_0, \beta), \]
\[ Z_n(u_0, \beta) = \max_{1 \leq j \leq n} \| G_{jh}(u_0, \beta) \|, \]
\[ V_n(u_0, \beta) = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0) G_{ih}(u_0, \beta) G_{ih}^*(u_0, \beta), \]

\[ V_n(u_0, \alpha, \beta) = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0) \frac{G_{ih}(u_0, \beta) G_{ih}^*(u_0, \beta)}{1 + \alpha \beta^r G_{ih}(u_0, \beta)}, \]
\[ B_n(u_0, \alpha, \beta) = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0) \frac{\partial G_{ih}(u_0, \beta)}{1 + \alpha \beta^r G_{ih}(u_0, \beta)}, \]
\[ C_n(u_0, \alpha, \beta) = \frac{1}{n} \sum_{i=1}^{n} \frac{K_h(U_i - u_0)}{(1 + \alpha \beta^r G_{ih}(u_0, \beta))^2} \frac{\partial G_{ih}(u_0, \beta)}{\partial \beta^r} G_{ih}(u_0, \beta), \]
\[ D_n(u_0, \alpha, \beta) = \psi^2 \frac{1}{n} \sum_{i=1}^{n} \frac{K_h(U_i - u_0)}{(1 + \alpha \beta^r G_{ih}(u_0, \beta))^2} \frac{\partial G_{ih}(u_0, \beta)}{\partial \beta^r}, \]
\[ E_n(u_0, \alpha, \beta) = \frac{1}{n} \sum_{i=1}^{n} \frac{K_h(U_i - u_0)}{(1 + \alpha \beta^r G_{ih}(u_0, \beta))^2} \frac{\partial G_{ih}(u_0, \beta)}{\partial \beta^r} \frac{\partial G_{ih}(u_0, \beta)^r}{\partial \beta}. \]

Here and hereafter the norm of a matrix \( W = (w_{ij}) \) is defined by \( \| W \| = \sqrt{\sum_{i,j} w_{ij}^2} \). Let \( r_0 \) denote an arbitrary positive constant. Let \( \Theta_0 \) be a compact subset of \( \mathbb{R}^p \) such that \( \beta_0 \) is an inner point of \( \Theta_0 \). Define

\[ \mathcal{F}_1 = \{ K((u - u_0)/h) G(u_0, \beta) : u_0 \in \Omega, \| \beta - \beta_0 \| \leq r_0 \}. \]
\[ \mathcal{F}_2 = \{ K((u - u_0)/h) G_{ih}(u_0, \beta) G_{ih}^*(u_0, \beta) : u_0 \in \Omega, \| \beta - \beta_0 \| \leq r_0 \}. \]
\[ \mathcal{F}_3 = \{ K((u - u_0)/h) \frac{\partial G_{ih}(u_0, \beta)}{\partial \beta^r} : u_0 \in \Omega, \beta \in \Theta_0 \}. \]

Let \( P_n \) denote the empirical distribution of \( \{(U_i, X_i, Y_i)\} \), and \( N(\delta, L_1(P_n), \mathcal{F}_j), \) \( j = 1, 2, 3 \) the covering numbers (see, e.g., Pollard, 1984, pp. 25 for the definition). We impose the following technical conditions:

(K0). \( K \) has the support \([-1, 1]\) and \( \max_t K(t) < \infty \)

(U0). The density of \( U \) is Lipschitz continuous and bounded from zero.

(A1). \( E[G(\varepsilon)|U] = 0 \). and \( \varepsilon \) is independent of \( X \) given \( U \).

(A2). There exist a constant \( \xi \geq 4 \) and a function \( F(Y, X) \) satisfying

\[ \sup_{|h| \leq 1, \| \beta - \beta_0 \| \leq \varepsilon_0} \| G(Y - \beta^r Z(X, t)) \| \| Z(X, t) \| \leq F(Y, X), \]
\[ \sup_{U} E[F(Y, X)^{\xi}|U = u] < \infty. \]

(A3). For \( 1 \leq k \leq k_0 \),

\[ \sup_{|\beta - \beta_0| \leq \varepsilon_0, u_0 \in \Omega, |t| \leq 1} E[G_{ih}(Y - \beta^r Z(X, t)) \| Z(X, t) \|^2|U = u_0 + th] = O(1). \]
(A4). There exist \(c_1(P_n)\) and some positive constant \(c_1\) such that \(E_{C_1}(P_n) \to c_1\) and
\[
N(\delta, L_1(P_n), F_1) \leq c_1(P_n)(h\delta)^{-w_1}.
\]

(A5). Uniformly for \(||\beta - \beta_0|| \to 0\) and \(h \to 0\),
\[
E\{G(Y - \beta^r Z(X, (U - u_0)/h)|U\} = O(h^2) + O(||\beta - \beta_0||).
\]

(A6). There exist \(c_2(P_n)\) and some positive constant \(c_2\) such that \(E_{C_2}(P_n) \to c_2\) and
\[
N(\delta, L_1(P_n), F_2) \leq c_2(P_n)(h\delta)^{-w_2}.
\]

(A7). \(\sup_{||\beta - \beta_0|| \leq r_0, u_0 \in \Omega, t \leq 1} E[G_Y(Y - \beta^r Z(X, t))||Z(X, t)||^4|U = u_0 + th] = O(1)\).

(A8). Uniformly for \(||\beta - \beta_0|| \to 0\) and \(h \to 0\),
\[
E\{G(Y - \beta^r Z(X, (U - u_0)/h)G^r(Y - \beta^r Z(X, (U - u_0)/h))|U\} = O(h^2) + O(||\beta - \beta_0||).
\]

(A9). \(V(u_0)\) and \(\Gamma(u_0)\) are Lipschitz continuous in \(u_0 \in \Omega\). Their minimum eigenvalues are uniformly positive in \(u_0 \in \Omega\).

(A10). For any \(\rho > 0\), there exists a constant \(c(\rho) > 0\) such that when \(h\) is small enough,
\[
\inf_{\beta \in \Theta_0, ||\beta - \beta_0|| \geq \rho} ||E_{K_h}(U - u_0)G_h(u_0, \beta)|| > c(\rho).
\]

For a positive sequence \(\rho_n \to 0\) and a small enough constant \(\rho_2\), as \(n \to \infty\),
\[
\inf_{\rho_1 \leq ||\beta - \beta_0|| \leq \rho_2} ||E_{K_h}(U - u_0)G_h(u_0, \beta)|| \geq \rho_n + O(h^2).
\]

(B1). There exist a constant \(\nu \geq 2\) and a function \(F_4(Y, X)\) such that
\[
\sup_u E[F_4^\nu(Y, X)|U = u| < \infty,
\]
\[
\sup_{u_0, \beta} \left| \frac{\partial G_h(u_0, \beta)}{\partial \beta^r} \right| ||I(|U - u_0| \leq h) \leq F_4(Y, X).
\]

(B2) For a constant \(c\),
\[
N(\delta, L_1(P_n), F_3) \leq c(h\delta)^{-w_3}.
\]

(B3). Uniformly for \(u_0 \in \Omega\) and \(||\beta - \beta_0|| \leq r_n = o(h^{1/2})\),
\[
E_{K_h}(U - u_0)\frac{\partial G_h(u_0, \beta)}{\partial \beta^r} = D(u_0) \otimes (S \otimes \Gamma(u_0)) + o(h^{1/2}).
\]

(B4). \(\sup_{||\beta - \beta_0|| \leq r_0, |t| \leq 1} E[|\frac{\partial^2 G_h(u_0, \beta)}{\partial \beta^r \beta^t}|^2|U = u_0 + th| < \infty\).

(B5). There exists function a \(F_5(y, x)\) such that
\[
\sup_u E[F_5^\nu(Y, X)|Y = u| < \infty,
\]
\[
\sup_{u_0 \in \Omega, ||\beta - \beta_0|| \leq r_0} \left| \frac{\partial^2 G_h(u_0, \beta)}{\partial \beta^r \beta^t} \right| |I(|U - u_0| \leq h) \leq F_5(Y, X).
\]

(B6). There exists a function \(F_6\) such that \(\sup_u E[F_6(\varepsilon, X)|X|^2|U = u| < \infty\), and that and for \(|U - u_0| \leq h\),
\[
|\partial G(\varepsilon + \frac{h^2}{2}, A^r(t)(u_0 + s(U - u_0))X(U - u_0)^2/h^2) + (\beta - \beta_0^r) Z(X, (U - u_0)/h))/\partial \varepsilon| \leq F_6(\varepsilon, X)
\]
uniformly for \(|s| \leq 1, ||\beta - \beta_0|| \leq r_0\), and \(u_0 \in \Omega\).
Remark 4.1 Suppressing dependence on $X$, we denote $Z(t) = Z(X, t)$. Suppose for some $r_0 > 0$, there exist integrable functions $F_j(Y,X), j = 1, 3$ such that

\[
\sup_{|\beta - \beta_0| \leq r_0, t} K'(t)||G(Y - \beta^* Z(t))|| ||Z(t)|| \leq F_1(Y,X),
\]

\[
\sup_{|\beta - \beta_0| \leq r_0, t} K(t)||\frac{\partial G(Y - \beta^* Z(t))}{\partial \epsilon}|| ||Z(t)||||Z'(t)|| + ||Z(t)|| \leq F_2(Y,X),
\]

\[
\sup_{|\beta - \beta_0| \leq r_0, t} K(t)||G(Y - \beta^* Z(t))|| ||Z'(t)|| \leq F_3(Y,X).
\]

Then for some positive constant $c$,

\[
||K((u - u_1)/h)G_h(u_1, \beta_1) - K((u - u_2)/h)G_h(u_2, \beta_2)||
\leq c(||F_1(Y,X) + F_2(Y,X) + F_3(Y,X)|||u_1 - u_2|/h + ||\beta_1 - \beta_2||).
\]

Thus the condition (A4) holds if $EF_j(Y,X) < \infty$, $j = 1, 2, 3$. The similar remarks can be made about the conditions (A6) and (B2).

As pointed out in Section 2, $EK_h(U - u_0)G_h(u_0, \beta_A) = 0, u_0 \in \Omega$ can be viewed as certain local estimating equations associated with the equations $E[G(Y - A(U)^* X)|U = u_0] = 0, u_0 \in \Omega$ as $A(u)$ is expanded around each $u_0$. In this sense, the first part of (A10) implies that when $\beta_A$ (coefficients of the approximation of $A$) is away from the true value $\beta_0$ (coefficients of the approximation of $A_0$), $||EK_h(U - u_0)G_h(u_0, \beta)||$ is away from 0. This is a little stronger than the requirement that $E[G(Y - A^*(U) X)|U] = 0$ if and only if $A$ is equal to the true value. The second part of (A10) is a local condition which says locally $||EK_h(U - u_0)G_h(u_0, \beta)||$ is bounded below by the norm of the linear function of $\beta$ near the true value $\beta_0$. For instance, assume the first component of $G$ is $Y - A^*(U) X$ and assume that $E[X X^*|U = u]$ is positive definite uniformly in $u$. Then we have

\[
||EK_h(U - u_0)G_h(u_0, \beta_A)|| \geq ||EK_h(U - u_0)[Y - \beta^*_A Z(X, (U - u_0)/h)] \otimes Z(X, (U - u_0)/h)||
\]

\[= O(h^2) + \beta_0 - \beta_A)^* \int K(t)E[Z(X,t)Z^*(X,t)|U = u_0 + th]f(u_0 + th)dt
\]

\[\geq c||\beta_0 - \beta_A|| + O(h^2),
\]

provided $h$ is small enough.

4.2 Proofs

Lemma 4.1 Under the conditions (K0), (U0), (A2)–(A4), if there exist some positive constants $b_0, b_1$ and $\eta < 1/2$ such that $b_0 \leq h^n \eta \leq b_1$, then there exists a sequence of positive constants $d_n \to 0$ such that

\[
A_n(u_0, \beta) = \frac{EK_h(U - u_0)G(Y - \beta^* Z(X, (U - u_0)/h)) \otimes Z(X, (U - u_0)/h)}{o_p(n^{-1/2} \wedge h^{1/2})d_n}.
\]

Furthermore, if the condition (A1) holds and $\eta > 1/(2\xi)$, then uniformly in $||\beta - \beta_0|| \leq r_n = o(n^{-1/2})d_n,$

\[
A_n(u_0, \beta) = o_p(n^{-1/2})d_n.
\]

Proof: Write

\[
A_n(u_0, \beta) = \frac{EK_h(U - u_0)G(Y - \beta^* Z(X, (U - u_0)/h)) \otimes Z(X, (U - u_0)/h)}{o_p(n^{-1/2} \wedge h^{1/2})d_n} + A_{n1}(u_0, \beta) + A_{n2}(u_0, \beta),
\]

14
where

\[
A_{n1}(u_0, \beta) = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0)G_h(u_0, \beta)I(F(Y_i, X_i) \leq M_n) \\
- EK_h(U - u_0)G_h(u_0, \beta)I(F(Y, X) \leq M_n),
\]

\[
A_{n2}(u_0, \beta) = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0)G_h(u_0, \beta)I(F(Y_i, X_i) > M_n) \\
- EK_h(U - u_0)G_h(u_0, \beta)I(F(Y, X) > M_n).
\]

Note that

\[
||EA_{n1}(u_0, \beta)|| \leq 2EK_h(U - u_0)||G_h(u_0, \beta)||I(F(Y, X) > M_n)|| \leq cM_n^{-1}. \tag{4.1}
\]

Consider the following empirical processes

\[
v_n(g) = n^{-1/2}M_n^{-1} \sum_{i=1}^{n} g(Y_i, X_i, u_0, \beta) - Eg(Y, X, u_0, \beta), \quad g \in F_1.
\]

Obviously, by the condition (A3),

\[
M_n^{-2}E||g(Y, X, u_0, \beta)||^2 \\
\leq c_n M_n^{-2} \sup_{\delta \in F_1} E_{Y|U=u_0+hG_h^2(Y - \beta^* Z(X, t))}|Z(X, t)||^2 \\
\leq O(hM_n^{-2}) = v.
\]

Now let \( M_n = n^\alpha, \delta_n = (h^{1/2} + n^{-1/\xi})(\log n)^{-1} \) and \( M = \delta_n n^{1/2} h M_n^{-1} \). Using Lemma 6.2 in Zhang and Gijbels (1999), we have

\[
P\{\sup_{g \in F_1} ||A_{n1}(u_0, \beta)|| \delta_n^{-1} > M_0\} = P \{\sup_{g \in F_1} ||v_n(g M_n^{-1})|| > M\} \\
\leq c_1 (n^{1/2}(\delta_n h M_n^{-1} M_n^{-1} |v|) \exp(-c_2 M^2/v) + c_2 v^{-\xi} \exp(-nu) \\
= O(\delta_n^{-\xi}) \exp(-c_3 M^{-2}/h M_n^{-2}) + c_2 O(h M_n^{-2})^{-\xi} \exp(-c_4 h M_n^{-2}). \tag{4.2}
\]

The last terms in (4.1) and (4.2) are \( o(\delta_n) \) and \( o(1) \) respectively if

\[
\begin{align*}
b_0 &\leq h_n \leq b_1, \quad nh^2/\log n \to \infty, \quad n^{1-2/\xi}h/\log n \to \infty, \\
nh M_n^{-2}/\log n &\to \infty, \quad M_n^{-\xi+1}\delta_n^{-1} \to 0.
\end{align*}
\]

The above requirements are fulfilled provided that \( s_0 > 0 \),

\[
\begin{align*}
b_0 &\leq h_n \leq b_1, \quad 0 < \eta < \min\{\frac{1}{2}, 1 - 2/\xi\}, \\
\max\{\eta/(2(\xi - 1)), 1/(\xi(\xi - 1))\} &< s_0 < (1 - \eta)/2.
\end{align*}
\]

These conditions are equivalent to

\[
\begin{align*}
b_0 &\leq h_n \leq b_1, \quad 0 < \eta < \min\{\frac{1}{2}, 1 - 2/\xi, 1 - 1/\xi, 1 - 2/(\xi(\xi - 1))\} = 1/2
\end{align*}
\]

since \( \xi \geq 4 \). The remaining part of the lemma is obvious because of the condition (A5). The proof is completed.
Lemma 4.2 Under the conditions (K0), (U0), (A2), (A6) and (A7), as $n \to \infty$, $b_0 \leq hn^\eta \leq b_1$, $0 < \eta < 1/2$, we have

$$V_n(u_0, \beta) = E K_h(U - u_0) G_h(u_0, \beta) G_h^*(u_0, \beta) + o_p(h^{1/2})$$

$$= V(u_0) \otimes (S \otimes \Gamma(u_0)) + o_p(h^{1/2}) + O(||\beta - \beta_0||).$$

Proof: The proof is similar to that of Lemma 1 if we replace $\xi$ there by $\xi/2$ and $M_n$ by $M_n^2$. In particular, note that

$$||E K_h(U - u_0) G_h(u_0, \beta) G_h^*(u_0, \beta)I(F(Y, X) > M_n)||$$

$$\leq \sup_u E|F(Y, X)^\xi |U = u| M_n^{2-\xi} = O(M_n^{2-\xi}).$$

Let $M_n = n^{\xi}$. Then a similar inequality to (4.2) holds if

$$b_0 \leq hn^\eta \leq b_1, \quad nh^2/\log n \to \infty,$$

$$nhM_n^{-4}/\log n \to \infty, \quad M_n^{2-\xi}h^{-1/2} \to 0.$$

These conditions are equivalent to

$$b_0 \leq hn^\eta \leq b_1, \quad 0 < \eta < \min\{1/2, 1 - 2/\xi\} = 1/2$$

due to the fact that $\xi \geq 4$. This completes the proof.

Lemma 4.3 Under the conditions (K0), (U0), (A1)~(A10), if there exists a sequence of positive constants $d_n \to 0$ such that as $n \to \infty$, $b_0 \leq hn^\eta \leq b_1$, $1/(2\xi) < \eta < 1/2$, then we have

$$\hat{\beta}(u_0) = \beta_0(u_0) + o_p(n^{-1/\xi} \land h^{1/2})d_n, \quad \alpha_n(u_0, \hat{\beta}(u_0)) = o_p(n^{-1/\xi} \land h^{1/2}).$$

Proof: First of all, by Lemma 4.1, we obtain

$$A_n(u_0, \beta_0) = o_p(n^{-1/\xi} \land h^{1/2})d_n. \quad (4.3)$$

Note that the condition (A2) implies

$$Z_n(u_0, \beta) = o_p(n^{1/\xi}) \quad (4.4)$$

uniformly in $u_0 \in \Omega$ and $||\beta - \beta_0|| \leq r_0$. Combining (4.4) with (4.3), Lemma 4.2 and the condition (A9), and

$$\alpha_n(u_0, \beta_0) = o_p(n^{-1/\xi} \land h^{1/2})d_n. \quad (4.5)$$

Set $\phi_n = (h^{1/2} \land n^{-1/\xi})d_n$, and let $u(u_0, \beta)$ satisfy

$$u(u_0, \beta)||EK_h(U - u_0) G_h(u_0, \beta)|| = E K_h(U - u_0) G_h(u_0, \beta).$$

Define

$$T_n(u_0, \beta) = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0) \log(1 + \phi_n u(u_0, \beta)^* G_{ih}(u_0, \beta)),$$

$$T_{n1}(u_0, \beta) = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0) \log(1 + \phi_n u(u_0, \beta)^* G_{ih}(u_0, \beta))I(||G_{ih}|| \leq n^{1/\xi}).$$
Then uniformly for $u_0$ and $\beta$,

$$T_{n1}(u_0, \beta) = \phi_n \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0)w(u_0, \beta)^T G\hat{u}_h(u_0, \beta)I(||G\hat{u}_h|| \leq n^{1/\ell})$$

$$- \frac{1}{2} \phi_n^2 |O(1)| \sum_{i=1}^{n} F(Y_i, X_i)^2$$

$$= \phi_n (u(u_0, \beta)^T EK_h(U - u_0)G_h(u_0, \beta) + o_p(\phi_n^2) + O_p(\phi_n^2). \quad (4.6)$$

Furthermore, by (4.3)–(4.6),

$$-T_n(u_0, \beta_0) \geq -|o_p(\phi_n^2)|.$$

Consequently, we have

$$P\{ \sup_{||\beta - \beta_0|| \geq \rho} (-T_n(u_0, \beta)) > -T_n(u_0, \beta_0), \text{ for some } u_0 \}$$

$$\leq P\{ \inf_{||\beta - \beta_0|| \geq \rho} \|EK_h(U - u_0)G_h(u_0, \beta)|| \leq |O_p(\phi_n)|, \text{ for some } u_0 \}$$

which together with the condition (A10) leads to

$$\hat{\beta}(u_0) - \beta_0 = O_p(\phi_n) = o_p(n^{-1/\ell} \wedge h^{1/2})d_n.$$

Invoking the argument of Owen (1988) and Lemma 4.1 again, we have

$$\alpha_n(u_0, \hat{\beta}(u_0)) = o_p(n^{-1/\ell} \wedge h^{1/2})$$

uniformly in $u_0$. This completes the proof.

**Lemma 4.4** Suppose for some positive constants $b_0$ and $b_1$, $b_0 \leq h \eta \leq b_1$, $0 < \eta < 1/2$. Then under the conditions (K0), (U0), (A2), (A6), (A7) and (A9), as $n \to \infty$, we have

$$V_n(u_0, \alpha, \beta) = V(u_0) \otimes (S \otimes \Gamma(u_0))(1 + o_p(h^{1/2}))$$

uniformly for $u_0 \in \Omega, ||\alpha|| + ||\beta - \beta_0|| \leq \alpha(n^{-1/\ell} \wedge h^{1/2})$.

**Proof:** Note that under the condition (A2), we have

$$\sup_{u_0 \in \Omega, ||\beta - \beta_0|| \leq \rho_0} Z_n(u_0, \beta) = o_p(n^{-1/\ell})$$

which together with Lemma 2.2 yields

$$V_n(u_0, \alpha, \beta) = V_n(u_0, \beta) + O_p(||\alpha||) \frac{(2 + o_p(1))}{(1 + o_p(1))} \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0)F(Y_i, X_i)^3$$

$$= V_n(u_0, \psi_2) + O_p(||\alpha||)$$

$$= V(u_0) \otimes (S \otimes \Gamma(u_0)) + o_p(h^{1/2}) + O_p(||\alpha||).$$

The proof is completed.
Lemma 4.5 Suppose there exist positive constants \( b_0, b_1 \) and \( \eta \) such that \( b_0 \leq \eta n^\eta \leq b_1, \ 0 < \eta < 1/2 \). Then under the conditions \((K0), (U0), (A2), (B1)\sim(B4)\), as \( n \to \infty \),

\[
B_n(u_0, \alpha, \beta) = D(u_0) \otimes (S \otimes \Gamma(u_0))(1 + o_p(h^{1/2}))
\]

uniformly for \( u_0 \in \Omega, \|\alpha\| + \|\beta - \beta_0\| \leq o(n^{-1/2}) \wedge h^{1/2} \).

Proof: Note that by the condition \((A2)\), we have

\[
\max \sup_{(u_0, \beta)} |\alpha^T G_{\beta \eta}(u_0, \beta)| = o_p(1)
\]

which with the condition \((B1)\) implies

\[
B_n(u_0, \alpha, \beta) = B_n(u_0, 0, \beta) + O_p(\|\alpha\|).
\]

It remains to prove

\[
B_n(u_0, 0, \beta) = D(u_0) \otimes (S \otimes \Gamma(u_0)) + o_p(h^{1/2}). \tag{4.7}
\]

To this end, we write

\[
B_n(u_0, 0, \beta) = B_{n1}(u_0, \beta) + B_{n2}(u_0, \beta) + E K_h\left(U - u_0, \frac{\partial G_h(u_0, \beta)}{\partial \beta^T}\right)
\]

where

\[
B_{n1}(u_0, \beta) = \frac{1}{n} \sum_{i=1}^{n} K_h(u_i - u_0) \frac{\partial G_h(u_0, \beta)}{\partial \beta^T} I(F_4(Y, X) \leq M_n)
\]

\[
- E K_h\left(U - u_0, \frac{\partial G_h(u_0, \beta)}{\partial \beta^T}\right) I(F_4(Y, X) \leq M_n),
\]

\[
B_{n2}(u_0, \beta) = \frac{1}{n} \sum_{i=1}^{n} K_h(u_i - u_0) \frac{\partial G_h(u_0, \beta)}{\partial \beta^T} I(F_4(Y, X) > M_n)
\]

\[
- E K_h\left(U - u_0, \frac{\partial G_h(u_0, \beta)}{\partial \beta^T}\right) I(F_4(Y, X) > M_n).
\]

Obviously,

\[
B_{n2}(u_0, \psi_2) = O_p(M_n^{1-\nu}),
\]

\[
E(K((U - u_0)/h)||\partial G_h(u_0, \beta)/\partial \beta^T||/M_n)^2 \leq O(h M_n^{-2}).
\]

Thus, a sufficient condition for \((4.7)\) is

\[
\begin{align*}
\quad b_0 & \leq \quad \eta n^\eta \leq b_1, \quad M_n = n^\eta, \\
\quad n h^2 / \log n \to \infty, \quad n h M_n^{-2} / \log n \to \infty, \quad M_n^{1-\nu} h^{-1/2} \to 0.
\end{align*}
\]

These are equivalent to

\[
\begin{align*}
b_0 & \leq \eta n^\eta \leq b_1, \quad 0 < \eta \leq \min\{1/2, 1 - 1/\nu\} = 1/2
\end{align*}
\]

by noting that \( \nu \geq 2 \).
Lemma 4.6 Under the conditions (K0), (U0), (A2), (B1), as \( h \to 0, nh \to \infty \),

\[
C_n(u_0, \alpha, \beta) = O_p(||\alpha||)
\]

uniformly for \( u_0 \in \Omega, ||\alpha|| + ||\beta - \beta_0|| \leq o(n^{-1/\ell} \wedge h^{1/2}) \).

Proof: Note that by the condition (A2) and \( ||\psi_2|| \leq o(n^{-1/\ell} \wedge h^{1/2}) \), we have

\[
\max_{i} \sup_{\beta_0} ||\alpha^T G_{ih}(u_0, \beta)|| = o_p(1).
\]

Thus

\[
||C_n(u_0, \alpha, \beta)|| \leq O_p(||\alpha||) \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0) F(Y_i, X_i) F(Y_i, X_i)
\]

\[
= O_p(||\alpha||)
\]

by the conditions (A2) and (B1). The proof is completed.

Lemma 4.7 Under the conditions (K0), (U0) and (B5), as \( h \to 0 \) and \( nh \to \infty \),

\[
D_n(u_0, \alpha, \beta) = O_p(||\alpha||)
\]

uniformly for \( u_0 \in \Omega, ||\alpha|| + ||\beta - \beta_0|| \leq o(n^{-1/\ell} \wedge h^{1/2}) \).

Proof: The proof is similar to the proof of the last lemma and hence is omitted.

Lemma 4.8 Under the conditions (K0), (U0) and (B5), as \( h \to 0, nh \to \infty \),

\[
E_n(u_0, \alpha, \beta) = O_p(||\alpha||^2)
\]

uniformly for \( u_0 \in \Omega, ||\alpha|| + ||\beta - \beta_0|| \leq o(n^{-1/\ell} \wedge h^{1/2}) \).

Proof: The proof is similar to the proof of the last lemma and thus omitted.

Proof of Theorem 1: First of all, using Lemma 4.3, we obtain

\[
\hat{\beta}(u_0) - \beta_0 = o_p(h^{1/2} \wedge n^{-1/\ell}), \quad \hat{\alpha}(u_0) = o_p(h^{1/2} \wedge n^{-1/\ell}).
\]

Furthermore, by the definition of \( \hat{\alpha} (= \hat{\alpha}(u_0)) \) and \( \hat{\beta} (= \hat{\beta}(u_0)) \), we have

\[
0 = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0) \frac{G_{ih}(u_0, \hat{\beta})}{1 + \hat{\alpha}^T G_{ih}(u_0, \hat{\beta})},
\]

\[
0 = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0) \frac{\hat{\alpha}^T \partial G_{ih}(u_0, \hat{\beta})/\partial \beta^*}{1 + \hat{\alpha}^T G_{ih}(u_0, \hat{\beta})}.
\]

Then the Taylor expansion, we have

\[
0 = A_n(u_0, \beta_0) - V_n(u_0, \alpha_{n1}, \beta_{n1}) \hat{\alpha} + (B_n(u_0, \alpha_{n1}, \beta_{n1}) - C_n(u_0, \alpha_{n1}, \beta_{n1}))(\hat{\beta} - \beta_0)
\]

\[
0 = \{B_n(u_0, \alpha_{n2}, \beta_{n2} - C_n(u_0, \alpha_{n2}, \beta_{n2}))\hat{\alpha} + (D_n(u_0, \alpha_{n2}, \beta_{n2}) - E_n(u_0, \alpha_{n2}, \beta_{n2}))(\hat{\beta} - \beta_0)
\]

19
where \( \alpha_{nj}, j = 1, 2 \) are between \( \hat{\alpha} \) and 0 and \( \beta_{nj}, j = 1, 2 \), are between \( \hat{\beta} \) and \( \beta_0 \). By using Lemmas 4.4 to 4.8, the above equations become

\[
-A_n(u_0, \beta_0) = -(1 + \alpha_p(h^{1/2}))V(u_0) \otimes (S \otimes \Gamma(u_0)) \hat{\alpha} + (\alpha_p(h^{1/2}) + D(u_0) \otimes (S \otimes \Gamma(u_0))))(\hat{\beta} - \beta_0),
\]

\[
0 = (\alpha_p(h^{1/2}) + D(u_0) \otimes (S \otimes \Gamma(u_0))) \hat{\alpha} + \alpha_p(h^{1/2})(\hat{\beta} - \beta_0).
\]

It follows that

\[
(\hat{\beta} - \beta_0) = -[(D(u_0)^*V^{-1}(u_0)D(u_0))^{-1}D(u_0)V^{-1}(u_0) \otimes (S^{-1} \otimes \Gamma(u_0))^{-1} + \alpha_p(h^{1/2})]A_n(u_0, \beta_0),
\]

\[
\hat{\alpha} = [V^{-1}(u_0) - V^{-1}(u_0)(D^r(u_0)V^{-1}(u_0)D(u_0))^{-1}D(u_0)D^r(u_0)V^{-1}(u_0) + \alpha_p(h^{1/2})]A_n(u_0, \beta_0).
\]

Observe that

\[
A_n(u_0, \beta) = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0)G(Y_i - A^r(U_i)X_i + \frac{1}{2}A^r(U_i^*)X_i(U_i - u_0)^2
\]

\[
+ (\beta - \beta_0)^*Z(X_i, (U_i - u_0)/h)) \otimes Z(X_i, (U_i - u_0)/h) + \frac{h^2}{2}O_p(1) + O_p(||\beta - \beta_0||)
\]

where the last equality follows from the condition (B6) (or \( A \) is linear). Now the proof can be completed by some simple calculations.

**Proof of Theorem 2:** Under the conditions of Theorem 1, we have \( h \to 0 \), and \( nh^{3/2} \to \infty \). Recall that given \( U, \varepsilon \) and \( X \) are independent by the condition (A1). By the Taylor expansion and Lemma 4.4, there are matrices \( V_n^*(U_j) \) such that as \( n \to \infty \), uniformly in \( U_j \),

\[
V_n^*(U_j) = V(U_j) \otimes (S \otimes \Gamma(U_j))(1 + \alpha_p(h^{1/2}))
\]

\[
\hat{\alpha}(U_j) = V_n^*(U_j)^{-1} \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - U_j)G \hat{h}(U_j, \hat{\beta}).
\]

The last two equalities lead to

\[
\hat{l}(G) = \sum_{j=1}^{n} \hat{\alpha}(U_j)^* \sum_{i=1}^{n} \frac{K_h(U_i - U_j)}{\sum_{m=1}^{n} K_h(U_m - U_j)} G \hat{h}(U_j, \hat{\beta})
\]

\[
- \frac{1}{2} \sum_{j=1}^{n} \hat{\alpha}(U_j)^*V_n(U_j, \hat{\alpha}(U_j), \hat{\beta}) \hat{\alpha}(U_j)
\]

\[
= \sum_{j=1}^{n} \hat{\alpha}(U_j)^*[\frac{1}{\sum_{m=1}^{n} K_h(U_m - U_j)} V_n^*(U_j) - \frac{1}{2} V_n(U_j, \hat{\alpha}(U_j, \hat{\beta}))] \hat{\alpha}(U_j)
\]

\[
= \frac{1}{2}(1 + \alpha_p(h^{1/2})) \sum_{j=1}^{n} f^{-1}(U_j) \hat{\alpha}(U_j)^*[V(U_j) \otimes (S \otimes \Gamma(u_0))] \hat{\alpha}(U_j),
\]

where \( 0 \leq s^* \leq 1 \), and \( V_n(u, \alpha, \beta) \) is defined at Subsection 4.1. Note that we draw out the factor \( 1 + \alpha_p(h^{1/2}) \) from the inside of the summation in (4.8) because the \( \alpha_p(h^{1/2}) \) is uniform with respect to \( U_j \), \( 1 \leq j \leq n \) and \( \hat{\alpha}(U_j)^*[V(U_j) \otimes (S \otimes \Gamma(U_j))] \hat{\alpha}(U_j) / \hat{\alpha}(U_j)^* \hat{\alpha}(U_j) \), \( 1 \leq j \leq n \) are bounded away from 0 and \( \infty \) (see Condition (A9)). It follows from the definition of \( C(u) \) in Subsection 3.1 that \( C(u)V(u)C(u) = C(u) \). Thus, combining
(4.8) and Theorem 1, we obtain
\[
l(G) = \left( \frac{1}{2} + o_p(h^{1/2}) \right) \sum_{j=1}^{n} \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - U_j)(C(U_j)G(\epsilon_i))^T
\]
\[
\otimes \left( \begin{array}{c}
\Gamma^{-1}(U_j)X_i \\
\mu_2^{-1}(U_i - U_j)\Gamma^{-1}(U_j)X_i / h
\end{array} \right)^T f(U_j)^{-1} [V(U_j) \otimes (S \otimes \Gamma(U_j))]^T
\]
\[
\times \frac{1}{n} \sum_{k=1}^{n} K_h(U_k - U_j)(C(U_j)G(\epsilon_k))^T \otimes \left( \begin{array}{c}
\Gamma^{-1}(U_j)X_k \\
\mu_2^{-1}(U_k - U_j)\Gamma^{-1}(U_j)X_k / h
\end{array} \right) + \zeta_n
\]
\[
= \left(1 + o_p(h^{1/2})\right) \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_j)K_h(U_k - U_j)f^{-1}(U_j)
\]
\[
\times G^T(\epsilon_i)C(U_j)G(\epsilon_k)(1 + \mu_2^{-1}(U_i - U_j)(U_k - U_j)/h^2)X_i^T\Gamma^{-1}(U_j)X_k + \zeta_n \tag{4.9}
\]

where \( \zeta_n = 0 \) when \( A_0 \) is linear, and otherwise \( \zeta_n = O_p(\sqrt{h}) \). The last term in (4.9) can be decomposed as follows:
\[
(1 + o_p(h^{1/2}))L(G) = T_{n11} + T_{n121} + T_{n122} + T_{n21} + T_{n22} + \zeta_n \tag{4.10}
\]

where
\[
T_{n11} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_j)^2f^{-1}(U_j)[G^T(\epsilon_i)C(U_j)G(\epsilon_i) - E[G^T(\epsilon_i)C(U_j)G(\epsilon_i)](U_i, U_j)]
\]
\[
\times (1 + \mu_2^{-1}(U_i - U_j)^2 / h^2)X_i^T\Gamma^{-1}(U_j)X_i,
\]
\[
T_{n121} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_j)^2E[G^T(\epsilon_i)C(U_j)G(\epsilon_i)](U_i, U_j)
\]
\[
\times (1 + \mu_2^{-1}(U_i - U_j)^2 / h^2)X_i^T\Gamma^{-1}(U_j)X_i - E[X_i^T\Gamma^{-1}(U_j)X_i(U_i, U_j)]f^{-1}(U_j),
\]
\[
T_{n122} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_j)^2E[G^T(\epsilon_i)C(U_j)G(\epsilon_i)](U_i, U_j)
\]
\[
\times (1 + \mu_2^{-1}(U_i - U_j)^2 / h^2)E[X_i^T\Gamma^{-1}(U_j)X_i(U_i, U_j)]f^{-1}(U_j),
\]
\[
T_{n21} = \frac{1}{n^2} \sum_{i \neq k \neq \{i, j\}} K_h(U_i - U_j)K_h(U_k - U_j)G^T(\epsilon_i)C(U_j)G(\epsilon_k)
\]
\[
\times (1 + (U_i - U_j)(U_k - U_j)\mu_2^{-1}/h)X_i^T\Gamma^{-1}(U_j)X_k,
\]
\[
T_{n22} = \frac{K(0)}{n^2h^2} \sum_{i \neq k} \{K((U_k - U_i)/h)G^T(\epsilon_i)C(U_j)G(\epsilon_k)X_i^T\Gamma^{-1}(U_i)X_kf^{-1}(U_i)
\]
\[
+ K((U_i - U_k)/h)G^T(\epsilon_i)C(U_k)G(\epsilon_k)X_i^T\Gamma^{-1}(U_k)X_kf^{-1}(U_k))
\]

Observe that as \( nh^{3/2} \to \infty, h \to 0, \)
\[
T_{n122} = \frac{K(0)^2}{(nh)^2} \sum_{i=1}^{n} \text{tr}(C(U_i)V(U_i))pf^{-2}(U_i)
\]
\[
\times \frac{1}{n^2} \sum_{i \neq j} K_h(U_i - U_j)^2\text{tr}(C(U_j)V(U_i))
\]
\[
\times (1 + \mu_2^{-1}(U_i - U_j)^2 / h^2)\text{tr}(\Gamma^{-1}(U_j)\Gamma(U_i))(f(U_i)f(U_j))^{-1}
\]
\[
= \frac{K(0)^2}{nh^2} [\text{tr}(C(U)V(U))/f^2(U) + O_p(n^{-1/2})] + \Psi_n \tag{4.11}
\]
\[
\approx o_p(h^{-1/2}) + \Psi_n
\]

21
where
\[
\Psi_n = \frac{1}{n^2} \sum_{i \neq j}^{N} K_h(U_i - U_j) \text{tr}(C(U_j)V(U_i)) \\
\times (1 + \mu_1^{-1}(U_i - U_j)^2/h^2) \text{tr}(\Gamma^{-1}(U_j)\Gamma(U_i))(f(U_i)f(U_j))^{-1},
\]
\[
= \frac{(k_0 - 1)p|\Omega|}{h} \int K^2(t)(1 + \mu_1^{-1}t^2)dt + o_p(h^{-1/2}).
\]
This is because
\[
E\Psi_n = (1 + O(h))^2 \frac{p}{h} \int K(t)^2(1 + \mu_1^{-1}t^2)dtE\text{tr}(C(U)V(U)f^{-1}(U)) \\
= \frac{p(k_0 - 1)}{h} (1 + O(h))|\Omega| \int K^2(t)(1 + \mu_1^{-1}t^2)dt,
\]
\[
\text{Var}(\Psi_n) \leq O(n^{-1}h^{-2}) = o(h^{-1}).
\]

By a similar argument, we have the following equalities
\[
T_{n121} = \frac{K(0)^2}{(nh)^2} \sum_{i=1}^{n} \text{tr}(C(U_i)V(U_i))(X_i^T \Gamma^{-1}(U_i)X_i - E[X_i^T \Gamma^{-1}(U_i)X_i])f^{-1}(U_i)
\]
\[
= \frac{1}{n^2} \sum_{i \neq j}^{N} K_h(U_i - U_j)^2 \text{tr}(C(U_j)V(U_j))(1 + \mu_1^{-1}(U_i - U_j)^2/h^2) \\
\times (X_i^T \Gamma^{-1}(U_j)X_i - E[X_i^T \Gamma^{-1}(U_j)X_i])f^{-1}(U_j)
\]
\[
= \frac{K(0)^2}{nh^2} - O_p(n^{-1/2}) + o_p(h^{-1/2}); \quad (4.12)
\]
\[
T_{n22} = O_p(h^{-1/2}); \quad (4.13)
\]
\[
T_{n21} = o_p(h^{-1/2}) + \frac{n-2}{n^2} \sum_{i \neq k} \Phi_i \Phi_k G(\epsilon_i)
\]
\[
= o_p(h^{-1/2}) + \frac{n-2}{n^2} \sum_{i \neq k} \Phi_i \Phi_k G(\epsilon_i)
\] \quad (4.14)

where \(\Phi_{ikh}\) can be found in Theorem 2 and the last equality follows from the Hoeffding's decomposition for the variance of U-statistics. Now (4.10)\text{–}(4.14) imply (3.2). (3.3) can be proved by a similar argument by showing that
\[
l(A_0|G) = (1 + O(h^{1/2})) \frac{1}{2n^2} \sum_{j=1}^{n} A_j(U_j, \beta_0)[V(U_j) \otimes (S \otimes \Gamma(U_j))]^{-1} A_n(U_j, \beta_0).
\]

The proof is completed.

**Proof of Theorem 3:** Invoking the asymptotic representations in Theorem 2, we need only to prove the asymptotic normality of \(T_n\). To this end, we first calculate the variance of \(T_n\),
\[
\text{Var}(T_n) = \frac{(2 + o(1))}{n(n-1)} E\{G^T(\epsilon_1)\Phi_{12h} G(\epsilon_2)\}^2
\]
\[
= \frac{(2 + o(1))}{n(n-1)} \text{tr}\left[ E\{\Phi_{12h} G(\epsilon_2)G^T(\epsilon_2)\Phi_{12h} G(\epsilon_1)G^T(\epsilon_1)\} \right]
\]
\[
= \frac{2(1 + O(h))}{n(n-1)} \text{tr}\left[ EK_h(U_2 - U_1)^2 C(U_1)G(\epsilon_2)G^T(\epsilon_2)C(U_1)G(\epsilon_2)G^T(\epsilon_1) \right]
\]
\[
\times X_i^T \Gamma^{-1}(U_1)X_2X_i^T \Gamma^{-1}(U_1)X_1\}
\]

22
\[ \begin{array}{l}
\frac{2(1 + O(h))}{n(n - 1)h} \text{tr}(E \int K^*(t)^2 dt C(U_1)V(U_2)C(U_1)\text{tr}(\Gamma^{-1}(U_1)\Gamma(U_2))(f(U_2)f(U_1))^{-1}) \\
= \frac{2(1 + O(h))}{n(n - 1)h} \text{tr}(E f^{-1}(U_1)\text{tr}(C(U_1)V(U_1))) \int K^*(t)^2 dt \\
= \frac{2(1 + O(h))}{n(n - 1)h} \frac{p(k_0 - 1)}{h} [\Omega] \int K^*(t)^2 dt.
\end{array} \] (4.15)

Let \( D_i = (\varepsilon_i, X_i, U_i), 1 \leq i \leq n, \) and \( \Pi_k \) be the \( \sigma \)-algebra generated by \( D_1, \ldots, D_k, 1 \leq k \leq n. \) Set \( \Phi_h(D_i, D_k) = G^r(\varepsilon_i)\Phi_h \Phi_h G(\varepsilon_k), \eta_{n1} = 0, \) and

\[ \eta_{nk} = E[T_n|\Pi_k] - E[T_n|\Pi_{k-1}]. \]

Then

\[ \eta_{nk} = \frac{2}{n(n - 1)} \sum_{j=1}^{k-1} \Phi_h(D_j, D_k), \quad 2 \leq k \leq n \]

and \( \{\eta_{nk}, \Pi_k\} \) is a sequence of martingale differences. By Theorem 4 of Shiryaev (1996, pp.543), it suffices to show

\[ \text{Var}^{-1}(T_n) \sum_{k=2}^{n} E[\eta_{nk}^2|\Pi_{k-1}] \rightarrow 1 \text{ in probability} \] (4.16)

and

\[ \text{Var}^{-2}(T_n) \sum_{k=1}^{n} E\eta_{nk}^4 \rightarrow 0. \] (4.17)

In the following, \( D = (\varepsilon, X, U) \) denotes a general random variable independent of \( D_i \) and \( D_k. \) To prove (4.16) and (4.17), we need the following equalities for \( i < j, \)

\[ E[\Phi_h(D_i, D_j)^2|D_j] = \frac{1}{h^2} \int K^*(t)^2 dt X_t^r \Gamma^{-1}(U_i)X_t G^r(\varepsilon_i)C(U_i)G(\varepsilon_i)^2 (1 + O(h)), \]

\[ E[\Phi_h(D_i, D)\Phi_h(D_j, D)|\Pi_{j-1}] = G(\varepsilon_i)^r E[K_h(U - U_i)K_h(U - U_k)C(U_i)V(U)C(U_j)X_t^r \Gamma^{-1}(U_j)X_t X_j^r]\]

\( \times \text{tr}(\Gamma^{-1}(U_k)XX^r \Gamma^{-1}(U_j)X_t X_t)\] \( (D_i, D_j)|G(\varepsilon_j)), \)

\[ E\eta_{nk}^2(D_i, D)\Phi_h(D_j, D) = \frac{1}{h^2}(1 + O(h)) \int \int K^*(t)^2 K^*(s)^2 dt ds E[(X_t^r \Gamma^{-1}(U_j)X_t)^2 \times G^r(\varepsilon_j)C(U_j)G(\varepsilon_j)^2]
\]

\[ \text{for} \quad \Phi_h(D_i, D_k) = O(1)(1 + O(h)) \frac{1}{h^3} \int K^*(t)^4 dt. \]

These are obvious by the assumption that \( \varepsilon \) and \( X \) are independent given \( U. \) Now with the above equalities, we can derive

\[ \sum_{k=2}^{n} E[\eta_{nk}^2|\Pi_{k-1}] = \sum_{k=2}^{n} \frac{4}{n^2(n - 1)^2} \left( \sum_{j=1}^{k-1} E[\Phi_h(D_j, D_k)^2|D_j] \right. \]

\[ + \sum_{i < j} E[\Phi_h(D_i, D_k)\Phi_h(D_j, D_k)|\Pi_{j-1}] \right) \]

\[ = \sum_{k=2}^{n} \frac{4}{n^2(n - 1)^2} \frac{1 + O(h)}{h} \int K^*(t)^2 dt X_t^r \Gamma^{-1}(U_i)X_t G^r(\varepsilon_i)C(U_i)G(\varepsilon_i) \]

23
\[
+ \sum_{k=2}^{n} \frac{8}{n^2(n-1)^2} \sum_{i<j}^{n-1} (n-j)E[\Phi_h(D_i, D)\Phi_h(D_j, D)\Phi(D_i, D_j)]
\]
\[
= (1 + O(h)) \frac{4}{n^2(n-1)^2} \int K^*(t)^2 dt \sum_{i=1}^{n-1} (n-i)X_i^T \Gamma^{-1}(U_i)X_i G^*(\epsilon_i)C(U_i)G(\epsilon_i) + \Upsilon_n
\]
\[
= (1 + o(1)) \frac{2}{n(n-1)} \int E[X_i^T \Gamma^{-1}(U_i)X_i | U_i]E[G^*(\epsilon_i)C(U_i)G(\epsilon_i)|U_i] + \Upsilon_n
\]
\[
= (1 + o(1)) Var(T_n) + \Upsilon_n
\] (4.18)

where
\[
\Upsilon_n = \frac{8}{n^2(n-1)^2} \sum_{i<j}^{n-1} (n-j)G(\epsilon_i)^2 E[K_h^*(U-U_i)K_h^*(U-U_h)C(U_i)C(U_h) C(U_i)C(U_h)]
\]
\[
\times \text{tr}(\Gamma^{-1}(U_h)XX^T \Gamma^{-1}(U_i)X_i X_i^T(U_i, U_h, X_i, X_h))(U_i, U_h, X_i, X_h)G(\epsilon_h).
\]

Note that
\[
E[\Upsilon_n] = \frac{64}{n^2(n-1)^2} \sum_{i<k}^{n-1} (n-k)E[G(\epsilon_i)^2 E[K_h^*(U-U_i)K_h^*(U-U_h)C(U_i)C(U_h) C(U_i)C(U_h)]
\]
\[
\times \text{tr}(\Gamma^{-1}(U_h)XX^T \Gamma^{-1}(U_i)X_i X_i^T(U_i, U_h, X_i, X_h))(U_i, U_h, X_i, X_h)G(\epsilon_h))^2
\]
\[
= O\left(\frac{(k_0 - 1)p}{(n-1)^4h^4}\right) \int [K^*(t) * K^*(t)]^2 dt
\]
\[
= O\left(\frac{1}{n^2} \text{Var}(T_n)\right),
\]

which implies \( \Upsilon_n = o_p(\text{Var}(T_n)) \) and where \( K^*(t) * K^*(t) \) is the convolution of \( K^*(t) \) with itself. Substituting the above equality into (4.18), we get (4.16). Analogously, (4.17) follows from the following calculations

\[
\sum_{k=2}^{n} E\eta_{nk}^2 = \frac{O(1)}{n^2(n-1)^2} \sum_{k=2}^{n} \sum_{i<j}^{k-1} \{O(\frac{1}{h^2}) + O(\frac{k-1}{h^2})\}
\]
\[
= \frac{O(1)}{n^2(n-1)^2} \left\{2 \sum_{i<j}^{n-1} (n-j)O(\frac{1}{h^2}) + O(\frac{1}{h^2}) \frac{n^2}{2}\right\}
\]
\[
= O(\text{Var}(T_n)^2) \frac{1}{(n-1)^2} \left\{O(n) + O(\frac{1}{h})\right\}.
\]

The proof is completed.

**Proof of Theorem 4:** The first part is similar to the proof of Theorem 3. The details are omitted. To show the second part, we recall that \( \Gamma(u_0) = E[XX^T | U = u_0]f(u_0) \) and write

\[
X_k = \begin{pmatrix} X_k^{(1)} \\ X_k^{(2)} \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}, \quad \text{and} \quad \Gamma_{11,2} = \Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}
\]

where \( X_k^{(1)} \) is \( p_1 \) dimensional, \( \Gamma_{11}, \Gamma_{12}, \Gamma_{21}, \Gamma_{22} \) are \( p_1 \times p_1, p_1 \times p_2, p_2 \times p_1 \) and \( p_2 \times p_2 \) matrices and \( p_2 = p - p_1 \). Following the same steps in the proof of Theorem 3, we first extend Theorem 1 as follows:

\[
\hat{\beta}_2(u_0) = \beta_2(u_0) + \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0) \begin{pmatrix} \Gamma_{22}^{-1}(u_0)X_i^{(2)} \\ \mu_2^{-1}\Gamma_{22}^{-1}(u_0)X_i^{(2)}(U_i - u_0)/h \end{pmatrix} \eta_i(u_0)(1 + o_p(h^{1/2})) + O_p(h^2),
\]

24
\[ \hat{a}^*(u_0) = \frac{1}{n} \sum_{i=1}^{n} K_h(U_i - u_0)((V^{-1}(u_0)G(\epsilon_i)) \odot \begin{pmatrix} \Gamma_1^{-1}(u_0)X_i \\ \mu_2^{-1}\Gamma_1^{-1}(u_0)X_i(U_i - u_0)/h \end{pmatrix} - \begin{pmatrix} 0 \\ \Gamma_2^{-1}(u_0)X_i^{(2)} \end{pmatrix}) \odot \begin{pmatrix} (1 + o_p(h^{1/2})) + O_p(h^2) \end{pmatrix}. \]

Then, by using the decomposition formula in Fan, Zhang and Zhang (2001), we have

\[ X_i^T \Gamma^{-1}(U_h)^{-1} X_k = \{X_i^{(1)}_i + \mu_2^{-1}\Gamma_2^{-1}(U_h)^{-1}X_i^{(2)} \} \Gamma_2^{-1}(U_h)^{-1}X_k - \frac{1}{h} V^{-1}(U_j)G(\epsilon_k + A(U_k)^T X_k). \]

The remaining part is very similar to the proof of Theorem 3. The details are omitted.

**Proof of Theorem 5:** The argument is similar to Fan, Zhang and Zhang (2001) but more tedious. For simplicity, we derive it heuristically. Write

\[ l(H_0 | G) = (1 + o_p(h^{1/2})) \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_j)K_h(U_k - U_j) \frac{1}{f(U_j)} G(\epsilon_i + A(U_i)^T X_i) \times V^{-1}(U_j)G(\epsilon_k + A(U_k)^T X_k)(1 + \mu_2^{-1}\frac{1}{h} V^{-1}(U_j)X_i^{(1)} \Gamma^{-1}(U_j)X_k - l_G) \]

\[ = (1 + o_p(h^{1/2}))(W_{n0} + W_{n1} + W_{n2} + W_{n3}) - l_G \]  

(4.19)

where

\[ W_{n0} = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_j)K_h(U_k - U_j)(1 + \mu_2^{-1}\frac{1}{h} V^{-1}(U_j)X_i^{(1)} \Gamma^{-1}(U_j)X_k) \]

\[ \times G(\epsilon_i) V^{-1}(U_j)G(\epsilon_k)X_i^{(2)} \]

\[ = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_j)K_h(U_k - U_j)(1 + \mu_2^{-1}\frac{1}{h} V^{-1}(U_j)X_i^{(1)} \Gamma^{-1}(U_j)X_k) \]

\[ \times G(\epsilon_i) V^{-1}(U_j) \frac{\partial G(\epsilon_i)X_i^{(1)} A(U_k)}{\partial \epsilon} \]

\[ W_{n2} = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_j)K_h(U_k - U_j)(1 + \mu_2^{-1}\frac{1}{h} V^{-1}(U_j)X_i^{(1)} \Gamma^{-1}(U_j)X_k) \]

\[ \times \frac{\partial G(\epsilon_i)X_i^{(1)} A(U_k)}{\partial \epsilon} \]

\[ W_{n3} = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_j)K_h(U_k - U_j)(1 + \mu_2^{-1}\frac{1}{h} V^{-1}(U_j)X_i^{(1)} \Gamma^{-1}(U_j)X_k) \]

\[ \times \frac{\partial G(\epsilon_i)X_i^{(1)} A(U_k)}{\partial \epsilon} \]

where \( \epsilon_i \) is between \( \epsilon_i + A(U_i)^T X_i \) and \( \epsilon_k \) is between \( \epsilon_k + A(U_k)^T X_k \). Under some regularity conditions,

\[ W_{n1} = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} G(\epsilon_i) V^{-1}(U_j) \]

\[ \{ \sum_{j=1}^{n} K_h(U_i - U_j)K_h(U_k - U_j)(1 + \mu_2^{-1}\frac{1}{h} V^{-1}(U_j)X_i^{(1)} \Gamma^{-1}(U_j)X_k) \]
\[ W_{n2} = W_{n1} \]

where \( W_{n1} \) is defined in (3.4). Similarly we write

\[
W_{n3} = W_{n31} + 2W_{n32} + W_{n33}
\]

where, when \( EA(U)^\tau XX^\tau A(U) = O(\frac{1}{nh^2}) \),

\[
W_{n31} = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_j)K_h(U_k - U_j)(1 + \mu_2 \frac{1}{h} \frac{U_i - U_j}{h} \frac{U_k - U_j}{h}) \frac{1}{f(U_j)} V^{-1}(U_j) \times \Xi_k A(U_i)^\tau X_i X_i^\tau \Gamma^{-1}(U_j) X_k X_k^\tau A(U_k)
\]

\[
= \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_k)V^{-1}(U_i)\Xi_k A(U_i)^\tau X_i X_i^\tau \Gamma^{-1}(U_j) X_k X_k^\tau A(U_k) + o_p(h^{-1/2})
\]

\[
= O\left(\frac{1}{nh^2}\right) + W_{n2}/2 + o_p(h^{-1/2}),
\]

\[
W_{n32} = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{n} E\left[\frac{\partial G(\epsilon_i)}{\partial \epsilon}_i \right] \frac{1}{f(U_j)} V^{-1}(U_j) \times E\left[\frac{\partial G(\epsilon_k)}{\partial \epsilon}_k \right] \frac{1}{f(U_j)} V^{-1}(U_j) \times X_i X_i^\tau \Gamma^{-1}(U_j) X_k X_k^\tau A(U_k)
\]

\[
= W_{n2} / 2 + o_p(h^{-1/2}),
\]

\[
W_{n33} = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{n} E\left[\frac{\partial G(\epsilon_i)}{\partial \epsilon}_i \right] \frac{1}{f(U_j)} V^{-1}(U_j) \times E\left[\frac{\partial G(\epsilon_k)}{\partial \epsilon}_k \right] \frac{1}{f(U_j)} V^{-1}(U_j) \times X_i X_i^\tau \Gamma^{-1}(U_j) X_k X_k^\tau A(U_k)
\]

\[
= O_p\left(\frac{1}{nh^2}\right) + \frac{n}{2} E\left[\frac{\partial G(\epsilon)}{\partial \epsilon}_U \right] V^{-1}(U) E\left[\frac{\partial G(\epsilon)}{\partial \epsilon}_A(U)^\tau XX^\tau A(U)(1 + o(1))\right]
\]

Recall that \( W_{n2}^{*} \) and \( W_{n3}^{*} \) are in (3.5) and (3.6) respectively. When \( EA(U)^\tau XX^\tau A(U) = O(\frac{1}{nh}) \), we have

\[
W_{n31} = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_k)V^{-1}(U_i)\Xi_k A(U_i)^\tau X_i X_i^\tau \Gamma^{-1}(U_j) X_k X_k^\tau A(U_k) + o_p(h^{-1/2})
\]

\[
= O\left(\frac{1}{nh^2}\right) + W_{n2}/2 + o_p(h^{-1/2}),
\]

\[
W_{n32} = W_{n2}^{*}/2 + o_p(h^{-1/2}),
\]

\[
W_{n33} = O_p\left(\frac{1}{nh^2}\right) + \frac{n}{2} E\left[\frac{\partial G(\epsilon)}{\partial \epsilon}_U \right] V^{-1}(U) E\left[\frac{\partial G(\epsilon)}{\partial \epsilon}_A(U)^\tau XX^\tau A(U)(1 + o(1))\right]
\]

Similarly, we have

\[
l_G = (1 + o_p(h^{-1/2})) \left[ \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_j)K_h(U_k - U_j)(1 + \mu_2 \frac{1}{h} \frac{U_i - U_j}{h} \frac{U_k - U_j}{h}) \right] \frac{1}{f(U_j)} G(\epsilon_i) C(U_j) G(\epsilon_k) + 2S_{n1} + S_{n3}
\]

(4.20)

where

\[
S_{n1} = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{n} K_h(U_i - U_j)K_h(U_k - U_j)(1 + \mu_2 \frac{1}{h} \frac{U_i - U_j}{h} \frac{U_k - U_j}{h}) \frac{1}{f(U_j)}
\]
\[
S_{n2} = \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{k=1}^{n} K_h(U_i - U_j)K_h(U_k - U_j)(1 + \mu_2^{-1} \frac{U_i - U_j}{h} \frac{U_k - U_j}{h}) \frac{1}{f(U_j)} \times \frac{\partial G(e_j^*)}{\partial e} C(U_j) \frac{\partial G(e_j^*)}{\partial e} \times A''(U_j) X_k X_k A''(U_j)(U_i - U_j)^2(U_k - U_j)^2
\]

where $U_j^*$ is between $U_k$ and $U_j$. Now the desired result follows from (4.19) and (4.20). This proves the theorem.

Department of Statistics, Chinese University of Hong Kong, Shatin, Hong Kong.
E-mail: jfan@u30b.stat.cuhk.edu.hk

EURANDOM, P.O.Box 513, 5600MB Eindhoven
and Institute of Systems Science, Academia Sinica, Beijing 100080.
E-mail: jzhang@euridice.tue.nl

References


