Report 2000-046

Sieve Empirical Likelihood Ratio Tests for Nonparametric Functions

Jianqing Fan and Jian Zhang

ISSN 1389-2355

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 χ^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.
AMS 1991 subject classification. 62G07, 62G10, 62J12.

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), Inglot 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

$$Y = a_1(U)X_1 + \cdots + a_p(U)X_p + \varepsilon, \tag{1.1}$$

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 χ^2-distribution (in a generalized sense) for testing the homogeneity

$$H_0 : a_1(\cdot) = \theta_1, \cdots, a_p(\cdot) = \theta_p \tag{1.2}$$

and for testing the significance of variables such as

$$H_0 : a_1(\cdot) = a_2(\cdot) = 0. \tag{1.3}$$

In other words, the generalized likelihood ratio statistic λ_n follows asymptotically a rescaled χ^2−distribution in the sense that $(2b_n)^{-1/2}(r_K\lambda_n - b_n) \xrightarrow{\mathcal{L}} N(0,1)$ for a sequence $b_n \to \infty$ and a constant r_K. We will use the notation $r_K\lambda_n \overset{a}{\sim} \chi^2_{b_n}$ to denote the result. The significance of the result is that the scale constant c_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 $a_3(\cdot), \cdots, a_p(\cdot)$ 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 ε 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 ε 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

$$E[G(\varepsilon)|U] = 0 \tag{1.4}$$

where $G = (G_1, \cdots, G_{k_o})^\tau$ 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 ε. In particular, when the conditional distribution of ε given U is symmetric about 0, we may choose a sequence of k_o grid points, say, $0 = s_o < s_1 < \cdots < s_{k_o}$ and take

$$G_k(\varepsilon) = I(\varepsilon \in [s_{k-1}, s_k]) - I(-\varepsilon \in [s_{k-1}, s_k]), \quad 1 \le k \le k_o \tag{1.5}$$

or a smoother version of the function G_k. Note that as $\max_{1\le k\le k_o}(s_k - s_{k-1}) \to 0$, $k_o \to \infty$, 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 ε 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 $a_{10}, \cdots, a_{p0}$. Our approach is to first construct the local linear estimator of the coefficient functions $a_1, \cdots, a_p$ 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 ε. 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 χ^2-distribution. The scaling constant and the degree of freedom are independent of the conditional distribution of ε 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):

$$Y = A^\tau(U)X + \varepsilon, \tag{2.1}$$

where Y is the response, $U \in \Omega \subset R^1$ (with Ω bounded) and $X \in R^p$ are covariates, ε is the stochastic error and $A(U) = (a_1(u), \cdots, a_p(u))$ 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

$$\begin{aligned} &Y \approx \beta_A(u_0)^T Z(X, (U - u_0)/h) + \varepsilon, \quad \text{for } U \approx u_0, \\ &E[G(Y - \beta_A(u_0)^\tau Z(X, (U - u_0)/h)|U = u] \approx 0, \quad \text{for } u \approx u_0 \end{aligned} \tag{2.2}$$

where $\beta_A(u_0) = (A^\tau(u_0), hA'^\tau(u_0))^\tau$ and $Z(X,t) = (X^\tau, tX^\tau)^\tau$. 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, ε 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

$$\mathbf{G}_{ih} = \mathbf{G}_{ih}(u_0, \beta) = G(Y_i - \beta^\tau Z(X_i, (U_i - u_0)/h)) \otimes Z(X_i, (U_i - u_0)/h)$$

with $\otimes$ being the Kronecker product, $\beta = (A^\tau, hB^\tau)^\tau$, $A = (a_1, \cdots, a_p)^\tau$ and $B = (b_1, \cdots, b_p)^\tau$. To see why we need an extra factor $Z(X_i, (U_i - u_0)/h)$ in the unconditional estimating function $\mathbf{G}_{ih}$, we let $G(\varepsilon) = \varepsilon$ temporally. 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 β_A. This leads to the following estimating equation

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

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 $\mathbf{G}_{ih}$.

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

$$l(\beta, u_0) = \sup\{\sum_{i=1}^n w_h(U_i, u_0) \log p_i : p_i \geq 0, 1 \leq i \leq n, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \mathbf{G}_{ih}(u_0, \beta) = 0\}$$

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

$$l(\beta, u_0) = \sum_{i=1}^n w_h(U_i, u_0) \log w_h(U_i, u_0) - \sum_{i=1}^n w_h(U_i, u_0) \log(1 + \alpha_n^\tau(u_0, \beta)\mathbf{G}_{ih}),$$

where $\alpha_n(u_0, \beta)$ satisfies

$$\sum_{i=1}^{n} w_h(U_i, u_0) \frac{\mathbf{G}_{ih}(u_0, \beta)}{1 + \alpha_n^\tau(u_0, \beta)\mathbf{G}_{ih}(u_0, \beta)} = 0. \tag{2.3}$$

Define the estimate of β 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 $\{E[G(\varepsilon)|U = U_j]\}_{1 \le j \le n}$ 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 $dP_{(Y,X)|U=U_j}(Y_j, X_j)$. 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 ε. 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:

$$\begin{aligned} l(G) &= l(\Theta|G) - l(\Theta|N), \\ &= \sum_{j=1}^{n} \sum_{i=1}^{n} w_h(U_i, U_j) \log(1 + \hat{\alpha}(U_j)^\tau \mathbf{G}_{ih}(U_j, \hat{\beta})) \end{aligned} \tag{2.6}$$

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:

$$H_{0h} : A(\cdot) = \theta.$$

More generally, we wish to test the composite nully hypothesis, which involves nuisance functions $A_2(\cdot)$:

$$H_{0u}: A_1 = A_{10} \qquad \longleftrightarrow \qquad H_{1u}: A_1 \neq A_{10} \tag{2.7}$$

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:

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

Let

$$\begin{aligned} \hat\beta_2(u_0) &= (\hat A_2^\tau, h\hat B_2^\tau)^\tau = \operatorname{argmax}_{\beta_2} l(\beta^*, u_0), \\ \hat\beta^*(u_0) &= (A_{10}^\tau(u_0), \hat A_2^\tau, hA_{10}'^\tau(u_0), h\hat B_2^\tau)^\tau \end{aligned}$$

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

$$0 = \frac{1}{n}\sum_{i=1}^n w_h(U_i, u_0)\frac{\mathbf{G}_{ih}(u_0, \hat\beta^*(u_0))}{1 + \hat\alpha^{*\tau}(u_0)\mathbf{G}_{ih}(u_0, \hat\beta^*(u_0))}.$$

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

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

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

$$H_{0s}: A = A_0 \qquad \longleftrightarrow \qquad H_{1s}: A \neq A_0 \tag{2.9}$$

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

$$\begin{aligned} l(H_{0s}|G) &= l(A_0|G) - l(\Theta|G) \\ &= \sum_{j=1}^n\sum_{i=1}^n w_h(U_i, U_j)\log(1 + \alpha_n(U_j, \beta_0)^\tau \mathbf{G}_{ih}(U_j, \beta_0)) - l(\Theta|G) \end{aligned} \tag{2.10}$$

where β_0 denotes β_{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

$$H_{0s}': A^* \equiv 0 \qquad \longleftrightarrow \qquad H_{1s}': A^* \not\equiv 0. \tag{2.11}$$

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:

$$H_0 : A \in \mathcal{A}_0 \qquad \longleftrightarrow \qquad A \notin \mathcal{A}_0, \tag{2.12}$$

where $\mathcal{A}_0$ is a set of functions. Let $l(\mathcal{A}_0|G)$ be the sieve empirical likelihood under the hypothesis H_0 in (2.12). Then, the sieve empirical likelihood ratio statistic is simply

$$\lambda_n = l(\mathcal{A}_0|G) - l(\Theta|G).$$

Let A_0' denote the true value of the parameter function A. Consider the fabricated testing problems with the simple null hypotheses:

$$H_0' : A = A_0', \qquad \longleftrightarrow \qquad H_1 : A \neq A_0' \tag{2.13}$$

and

$$H_0' : A = A_0', \qquad \longleftrightarrow \qquad H_1' : A \in \mathcal{A}_0. \tag{2.14}$$

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 $\lambda^*(A_0'|G) = l(A_0'|G) - l(\mathcal{A}_0|G)$ is the sieve empirical ratio test for the problem (2.14). Thus, the asymptotic representation of λ_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 $\hat{\beta}$ 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 $l(H_{0s}|G)$ 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

$$\begin{aligned}
D(u_0) &= -E[\frac{\partial G(\epsilon)}{\partial \varepsilon}|U = u_0], \quad V(u_0) = E[G(\varepsilon)G^\tau(\varepsilon)|U = u_0], \\
\Gamma(u_0) &= E[XX^\tau|U = u_0]f(u_0), \quad S = \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}, \quad \mu_2 = \int t^2K(t)dt. \\
\eta_i(u_0) &= -\{D(u_0)^\tau V(u_0)^{-1}D(u_0)\}^{-1}D^\tau(u_0)V^{-1}(u_0)G(\varepsilon_i), \\
C(u_0) &= V^{-1}(u_0) - V^{-1}(u_0)(D^\tau(u_0)V^{-1}(u_0)D(u_0))^{-1}D(u_0)D^\tau(u_0)V^{-1}(u_0), \\
\varepsilon_i &= Y_i - A^\tau(U_i)X_i.
\end{aligned}$$

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 $b_0 \le hn^{\eta} \le b_1$, then uniformly for $u_0 \in \Omega$,*

$$\begin{aligned}
\hat{\beta}(u_0) &= \beta(u_0) + \frac{1}{n}\sum_{i=1}^{n} K_h(U_i - u_0)\begin{pmatrix} \Gamma^{-1}(u_0)X_i \\ \mu_2^{-1}\Gamma^{-1}(u_0)X_i(U_i - u_0)/h \end{pmatrix}\eta_i(u_0)(1 + o_p(h^{1/2})) + O_p(h^2), \\
\hat{\alpha}(u_0) &= \frac{1}{n}\sum_{i=1}^{n} K_h(U_i - u_0)\{C(u_0)G(\varepsilon_i)\} \otimes \begin{pmatrix} \Gamma^{-1}(u_0)X_i \\ \mu_2^{-1}\Gamma^{-1}(u_0)X_i(U_i - u_0)/h \end{pmatrix}(1 + o_p(h^{1/2})) + O_p(h^2).
\end{aligned}$$

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

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

The asymptotic normality of the local 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 $E[G(\varepsilon - t)|U = u_0]$ is twice continuously differentiable in t. In this case $D(u_0)$ should be replaced by $-\{\partial E[G(\varepsilon - t)|U = u_0]/\partial t\}|_{t=0}$. Similar to Liu and Zhang (2000), we can show that the asymptotic efficiency of $\hat{A}(u_0)$ is increasing in $D(u_0)^{\tau}V(u_0)^{-1}D(u_0)$. In particular, in the setting of symmetric location model mentioned in Section 1, we can find a squence of G, say $\{G^{(k)}\}$ such that the corresponding $\hat{A}(u_0)$ is asymptotically adaptive to the unknown conditional density of ε given $U = u_0$. In practice, to save the computational effort, we prefer to choosing 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 $l(H_{0s}|G)$. The results indicate that they admit a generalized quadratic form. To facilitate the notation, the following notation is introduced. Let

$$\begin{aligned}
\phi_{ikh}(U) &= K_h(U_i - U)K_h(U_k - U)C(U)(1 + (U_i - U)(U_k - U)\mu_2^{-1}h^{-2})X_i^{\tau}\Gamma^{-1}(U)X_k f^{-1}(U), \\
K^*(s) &= \int K(t)K(s+t)(1 + t(s+t)\mu_2^{-1})dt, \qquad (3.1) \\
\Phi_{ikh} &= E[\phi_{ikh}(U)|(U_i, U_k, X_i, X_k)] \\
&= K_h^*(U_k - U_i)C(U_i)X_i^{\tau}\Gamma^{-1}(U_i)X_k(1 + O_p(h)), \\
T_n &= \frac{1}{n(n-1)}\sum_{i \ne k} G^{\tau}(\varepsilon_i)\Phi_{ikh}G(\varepsilon_k).
\end{aligned}$$

Similarly, we define

$$\begin{aligned}
q_{ikh}(U) &= K_h(U_i - U)K_h(U_k - U)V^{-1}(U)X_i^{\tau}\Gamma^{-1}(U)X_k\{1 + (U_i - U)(U_k - U)\mu_2^{-1}h^{-2}\}f^{-1}(U), \\
Q_{ikh} &= E[q_{ikh}(U)|(U_i, U_k, X_i, X_k)], \\
T_n^* &= \frac{1}{n(n-1)}\sum_{i \ne k} G^{\tau}(\varepsilon_i)\left(Q_{ikh} - \Phi_{ikh}\right)G(\varepsilon_k).
\end{aligned}$$

Then, we have the following result.

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

$$2l(G) = \frac{(k_o - 1)p|\Omega|}{h}\int K^2(t)(1 + t^2\mu_2^{-1})dt + (1 + o_p(h^{1/2}))nT_n + o_p(h^{-1/2}); \qquad (3.2)$$

and under H_{0s}, if A_0 is linear or $nh^{9/2} \to 0$, then

$$2l(H_{0s}|G) = \frac{p|\Omega|}{h}\int K^2(t)(1+t^2\mu_2^{-1})dt + (1+o_p(h^{1/2}))nT_n^* + o_p(h^{-1/2}), \quad (3.3)$$

where $|\Omega|$ is the length of the support Ω of the density f.

Note that if there are no components in A, then under H_{0G} 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 $l(H_{0s}|G)$. Like in the parametric case for the stochastic error ε (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 χ^2−distributed and their degrees of freedom are independent of the nuisance parameters such as A, G, and the distribution of ε.

Theorem 3 *Under H_{0G} and the conditions of Theorem 1, for $k_0 > 1$, we have $r_K l_G \overset{a}{\sim} \chi^2_{b_n}$ with*

$$r_K = \frac{2K^*(0)}{\int K^*(s)^2 ds}, \quad b_n = \frac{(k_o-1)p|\Omega|c_K}{h},$$

where $K^*(s)$ is defined in (3.1), $c_K = K^*(0)^2/\int K^*(s)^2 ds$. For $k_0 = 1$, we have $r_K l_G = o_p(1)$.

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*

$$|K^*(s)| \le \int K(t)K(s+t)dt + \mu_2^{-1}\int |t|K(t)|s+t|K(s+t)dt \le K^*(0).$$

Thus when $K^*(s) \ge 0, s \in [-1,1], r_K \ge 2$. In particular, when K is the uniform kernel function, $r_K = 2.8176$ and $c_K = 1.0566$; when K is the Epanechnikov kernel function, $r_K = 2.5154$ and $c_K = 1.2936$.

The next theorem presents the asymptotic null distribution of $l(H_{0s}|G)$.

Theorem 4 *Suppose that the conditions of Theorem 1 hold. Then under H_{0s}, $r_K l(H_{0s}|G) \overset{a}{\sim} \chi^2_{b_n^*}$; and under H_{0u}, if $nh^{9/2} \to 0$, then $r_K l(H_{0u}|G) \overset{a}{\sim} \chi^2_{b_{n2}^*}$ where $b_n^* = p|\Omega|c_K/h$ and $b_{n2}^* = p_1|\Omega|c_K/h$ with c_K and r_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 ε 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 ε is essentially known. In particular, if the variance is heteroscedastic with $\text{var}(\varepsilon|U) = \sigma^2(U)$, 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

$$W_{1n}^* = \frac{1}{n}\sum_{i\neq k}^{n} K_h^*(U_i - U_k)G(\varepsilon_i)V^{-1}(U_k)X_i^\tau\Gamma^{-1}(U_k)X_kA(U_k)^\tau X_k\frac{\partial G(\varepsilon_k)}{\partial\varepsilon}, \tag{3.4}$$

$$\begin{aligned} \Xi_i &= \frac{\partial G(\varepsilon_i)}{\partial\varepsilon} - E[\frac{\partial G(\varepsilon_i)}{\partial\varepsilon}|U_i], \\ W_{2n}^* &= \frac{1}{n}\sum_{i\neq k}^{n} K_h^*(U_i - U_k)\Xi_i^\tau V^{-1}(U_i)\Xi_k A(U_i)^\tau X_iX_i^\tau\Gamma^{-1}(U_k)X_kX_k^\tau A(U_k), \end{aligned} \tag{3.5}$$

$$W_{3n}^* = \frac{1}{n}\sum_{i\neq k}^{n} K_h^*(U_i - U_k)\Xi_i^\tau V^{-1}(U_k)E[\frac{\partial G(\varepsilon_k)}{\partial\varepsilon}|U_k]A(U_i)X_iX_i^\tau\Gamma^{-1}(U_k)X_kX_k^\tau A(U_k). \tag{3.6}$$

Then, following the same arguments used in Fan, Zhang and Zhang (2001), we can derive the asymptotic power $l(H_{0s}|G)$ 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 $nhEA(U)^\tau XX^\tau A(U) = O(1)$, $\max_u ||A(u)|| \to 0$ 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,*

$$\begin{aligned} 2l(H_{0s}|G) &= \frac{p|\Omega|}{h}K^*(0) + nE\{D(U)^\tau V^{-1}(U)D(U)A(U)^\tau XX^\tau A(U)\}(1+o(1)) \\ &\quad -\frac{nh^4}{4}E\{D(U)^\tau C(U)D(U)A''(U)^\tau XX^\tau A''(U)\} \\ &\quad \times\int\int t^2(s+t)^2K(t)K(s+t)(1+\mu_2^{-1}t(s+t))dtds(1+o(1)) \\ &\quad +(1+o_p(h^{1/2}))\{T_n^* + 2W_{1n}^* + W_{2n}^* + 2W_{3n}^*\} + o_p(h^{-1/2}) \end{aligned}$$

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 $h = c_* n^{-2/9}$ for some constant c_*. This rate is optimal in the ordinary nonparametric regression setting.

4 Technical conditions and proofs

4.1 Technical conditions

Define

$$
\begin{aligned}
A_n(u_0,\beta) &= \frac{1}{n}\sum_{i=1}^n K_h(U_i-u_0)\mathbf{G}_{ih}(u_0,\beta),\\
Z_n(u_0,\beta) &= \max_{1\le j\le n}||\mathbf{G}_{jh}(u_0,\beta)||,\\
V_n(u_0,\beta) &= \frac{1}{n}\sum_{i=1}^n K_h(U_i-u_0)\mathbf{G}_{ih}(u_0,\beta)\mathbf{G}_{ih}^{\tau}(u_0,\beta),
\end{aligned}
$$

$$
\begin{aligned}
V_n(u_0,\alpha,\beta) &= \frac{1}{n}\sum_{i=1}^n K_h(U_i-u_0)\frac{\mathbf{G}_{ih}(u_0,\psi_2)\mathbf{G}_{ih}^{\tau}(u_0,\beta)}{1+\alpha^{\tau}\mathbf{G}_{ih}(u_0,\beta)},\\
B_n(u_0,\alpha,\beta) &= \frac{1}{n}\sum_{i=1}^n \frac{K_h(U_i-u_0)}{1+\alpha^{\tau}\mathbf{G}_{ih}(u_0,\beta)}\frac{\partial\mathbf{G}_{ih}(u_0,\beta)}{\partial\beta^{\tau}},\\
C_n(u_0,\alpha,\beta) &= \frac{1}{n}\sum_{i=1}^n \frac{K_h(U_i-u_0)}{(1+\alpha^{\tau}\mathbf{G}_{ih}(u_0,\beta))^2}\frac{\partial\mathbf{G}_{ih}(u_0,\beta)}{\partial\beta^{\tau}}\alpha\mathbf{G}_{ih}^{\tau}(u_0,\beta),\\
D_n(u_0,\alpha,\beta) &= \psi_1^{\tau}\frac{1}{n}\sum_{i=1}^n \frac{K_h(U_i-u_0)}{1+\alpha^{\tau}\mathbf{G}_{ih}(u_0,\beta)}\frac{\partial^2\mathbf{G}_{ih}(u_0,\beta)}{\partial\beta\partial\beta^{\tau}},\\
E_n(u_0,\alpha,\beta) &= \frac{1}{n}\sum_{i=1}^n \frac{K_h(U_i-u_0)}{(1+\alpha^{\tau}\mathbf{G}_{ih}(u_0,\beta))^2}\frac{\partial\mathbf{G}_{ih}(u_0,\beta)}{\partial\beta^{\tau}}\alpha\beta^{\tau}\frac{\partial\mathbf{G}_{ih}(u_0,\beta)^{\tau}}{\partial\beta}.
\end{aligned}
$$

Here and hereafter the norm of a metrix $W=(w_{ij})$ is defined by $||W||=\sqrt{\sum_{i,j}^n w_{ij}^2}$. Let r_0 denote an arbitrary positive constant. Let Θ_0 be a compact subset of R^{2p} such that β_0 is an inner point of Θ_0. Define

$$\mathcal{F}_1=\{K((u-u_0)/h)\mathbf{G}(u_0,\beta):u_0\in\Omega,||\beta-\beta_0||\le r_0\}.$$

$$\mathcal{F}_2=\{K((u-u_0)/h)\mathbf{G}_h(u_0,\beta)\mathbf{G}_h^{\tau}(u_0,\beta):u_0\in\Omega,||\beta-\beta_0||\le r_0\}.$$

$$\mathcal{F}_3=\{K((u-u_0)/h)\frac{\partial\mathbf{G}_h(u_0,\beta)}{\partial\beta^{\tau}}:u_0\in\Omega,\beta\in\Theta_0\}$$

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 ε is independent of X given U.

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

$$
\begin{aligned}
&\sup_{|t|\le 1,||\beta-\beta_0||\le\delta_0}||G(Y-\beta^{\tau}Z(X,t))||\,||Z(X,t)||\le F(Y,X),\\
&\sup_u E[F(Y,X)^{\xi}|U=u]<\infty.
\end{aligned}
$$

(A3). For $1\le k\le k_0$,

$$\sup_{||\beta-\beta_0||\le r_0,u_0\in\Omega,|t|\le 1} E[G_k^2(Y-\beta^{\tau}Z(X,t))||Z(X,t)||^2|U=u_0+th]=O(1).$$

(A4). There exist $c_1(P_n)$ and some positive constant c_1 such that $Ec_1(P_n) \to c_1$ and

$$N(\delta, L_1(P_n), \mathcal{F}_1) \le c_1(P_n)(h\delta)^{-w_1}.$$

(A5). Uniformly for $||\beta - \beta_0|| \to 0$ and $h \to 0$,

$$E\{G(Y - \beta^\tau Z(X, (U - u_0)/h)|U\} = O(h^2) + O(||\beta - \beta_0||).$$

(A6). There exist $c_2(P_n)$ and some positive constant c_2 such that $Ec_2(P_n) \to c_2$ and

$$N(\delta, L_1(P_n), \mathcal{F}_2) \le c_2(P_n)(h\delta)^{-w_2}.$$

(A7). $\sup_{||\beta-\beta_0||\le r_0, u_0\in\Omega |t|\le 1} E[G_k^4(Y - \beta^\tau Z(X,t))||Z(X,t)||^4|U = u_0 + th] = O(1)$.
(A8). Uniformly for $||\beta - \beta_0|| \to 0$ and $h \to 0$,

$$E\{G(Y - \beta^\tau Z(X, (U - u_0)/h)G^\tau(Y - \beta^\tau Z(X, (U - u_0)/h))|U\} = O(h^2) + O(||\beta - \beta_0||).$$

(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,

$$\inf_{\beta\in\Theta_0, ||\beta-\beta_0||\ge\rho} ||EK_h(U - u_0)\mathbf{G}_h(u_0, \beta)|| > c(\rho).$$

For a positive sequence $\rho_{n1} \to 0$ and a small enough constant ρ_2, as $n \to \infty$,

$$\inf_{\rho_{n1}\le||\beta-\beta_0||\le\rho_2} ||EK_h(U - u_0)\mathbf{G}_h(u_0, \beta)|| \ge \rho_{n1} + O(h^2).$$

(B1). There exist a constant $\nu \ge 2$ and a function $F_4(Y, X)$ such that

$$\sup_u E[F_4^\nu(Y, X)|U = u] < \infty,$$
$$\sup_{u_0,\beta} ||\frac{\partial \mathbf{G}_h(u_0, \beta)}{\partial \beta^\tau}||I(|U - u_0| \le h) \le F_4(Y, X).$$

(B2) For a constant c,

$$N(\delta, L_1(P_n), \mathcal{F}_3) \le c(h\delta)^{-w_3}.$$

(B3). Uniformly for $u_0 \in \Omega$ and $||\beta - \beta_0|| \le r_n = o(h^{1/2})$,

$$EK_h(U - u_0)\frac{\partial \mathbf{G}_h(u_0, \beta)}{\partial \beta^\tau} = D(u_0) \otimes (S \otimes \Gamma(u_0)) + o(h^{1/2}).$$

(B4). $\sup_{||\beta-\beta_0||\le r_0, |t|\le 1} E[||\partial \mathbf{G}_h(u_0, \beta)/\partial\beta^\tau||^2|U = u_0 + th] < \infty$.
(B5). There exists function a $F_5(y, x)$ such that

$$\sup_u E[F_5^2(Y, X)|Y = u] < \infty,$$
$$\sup_{u_0\in\Omega, ||\beta-\beta_0||\le r_0} ||\frac{\partial^2 \mathbf{G}_h(u_0, \beta)}{\partial\beta\partial\beta^\tau}||I(|U - u_0| \le h) \le F_5(Y, X).$$

(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 $|U - u_0| \le h$,

$$||\partial G(\varepsilon + \frac{h^2}{2}A''^\tau(u_0 + s(U - u_0))X(U - u_0)^2/h^2) + (\beta - \beta_0)^\tau Z(X, (U - u_0)/h))/\partial\varepsilon|| \le F_6(\varepsilon, X)$$

uniformly for $|s| \le 1$, $||\beta - \beta_0|| \le r_0$, 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*

$$\sup_{||\beta-\beta_0||\le r_0,t} K'(t)||G(Y-\beta^\tau Z(t))||\ ||Z(t)|| \le F_1(Y,X),$$
$$\sup_{||\beta-\beta_0||\le r_0,t} K(t)||\frac{\partial G(Y-\beta^\tau Z(t))}{\partial \varepsilon}||\ ||Z(t)||(||Z'(t)|| + ||Z(t)||) \le F_2(Y,X),$$
$$\sup_{||\beta-\beta_0||\le r_0,t} K(t)||G(Y-\beta^\tau Z(t))||\ ||Z'(t)|| \le F_3(Y,X).$$

Then for some positive constant c,

$$\begin{aligned}&||K((u-u_1)/h)\mathbf{G}_h(u_1,\beta_1) - K((u-u_2)/h)\mathbf{G}_h(u_2,\beta_2)||\\ &\quad\le c\{|F_1(Y,X) + F_2(Y,X) + F_3(Y,X)\}\{|u_1-u_2|/h + ||\beta_1-\beta_2||\}.\end{aligned}$$

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, $EK_h(U-u_0)\mathbf{G}_h(u_0,\beta_A) = 0, u_0 \in \Omega$ can be viewed as certain local estimating equations associated with the equations $E[G(Y-A(U)^\tau X)|U = u_0] = 0, u_0 \in \Omega$ as $A(u)$ is expanded around each u_0. In this sense, the first part of (A10) implies that when β_A (coefficients of the approximation of A) is away from the true value β_0 (coefficients of the approximation of A_0), $||EK_h(U-u_0)\mathbf{G}_h(u_0,\beta)||$ is away from 0. This is a little stronger than the requirement that $E[G(Y-A^\tau(U)X)|U] = 0$ if and only if A is equal to the true value. The second part of (A10) is a local condition which says locally $||EK_h(U-u_0)\mathbf{G}_h(u_0,\beta)||$ is bounded below by the norm of the linear function of β near the true value β_0. For instance, assume the first component of G is $Y - A^\tau(U)X$ and assume that $E[XX^\tau|U = u]$ is positive definite uniformly in u. Then we have

$$\begin{aligned}&||EK_h(U-u_0)\mathbf{G}_h(u_0,\beta_A)|| \ge ||EK_h(U-u_0)[Y-\beta_A^\tau Z(X,(U-u_0)/h)] \otimes Z(X,(U-u_0)/h)||\\ &= O(h^2) + (\beta_0-\beta_A)^\tau \int K(t)E[Z(X,t)Z^\tau(X,t)|U = u_0+th]f(u_0+th)dt\\ &\ge c||\beta_0-\beta_A|| + O(h^2),\end{aligned}$$

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 $b_0 \le hn^\eta \le b_1$, then there exists a sequence of positive constants $d_n \to 0$ such that*

$$A_n(u_0,\beta) = EK_h(U-u_0)G(Y-\beta^\tau Z(X,(U-u_0)/h)) \otimes Z(X,(U-u_0)/h) + o_p(n^{-1/\xi} \wedge h^{1/2})d_n.$$

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

$$A_n(u_0,\beta) = o_p(n^{-1/\xi})d_n.$$

Proof: Write

$$\begin{aligned}A_n(u_0,\beta) &= EK_h(U-u_0)G(Y-\beta^\tau Z(X,(U-u_0)/h) \otimes Z(X,(U-u_0)/h)\\ &\quad +A_{n1}(u_0,\beta) + A_{n2}(u_0,\beta),\end{aligned}$$

where

$$\begin{aligned}
A_{n1}(u_0,\beta) &= \frac{1}{n}\sum_{i=1}^{n} K_h(U_i-u_0)\mathbf{G}_{ih}(u_0,\beta)I(F(Y_i,X_i)\le M_n) \\
&\quad -EK_h(U-u_0)\mathbf{G}_h(u_0,\beta)I(F(Y,X)\le M_n), \\
A_{n2}(u_0,\beta) &= \frac{1}{n}\sum_{i=1}^{n} K_h(U_i-u_0)\mathbf{G}_{ih}(u_0,\beta)I(F(Y_i,X_i)> M_n) \\
&\quad -EK_h(U-u_0)\mathbf{G}_h(u_0,\beta)I(F(Y,X)> M_n).
\end{aligned}$$

Note that

$$||EA_{n1}(u_0,\beta)|| \le 2EK_h(U-u_0)||\mathbf{G}_h(u_0,\beta)||I(F(Y,X)>M_n)|| \le cM_n^{1-\xi}. \tag{4.1}$$

Consider the following empirical processes

$$v_n(g) = n^{-1/2}M_n^{-1}\sum_{i=1}^{n}(g(Y_i,X_i,u_0,\beta)-Eg(Y,X,u_0,\beta)), \quad g\in\mathcal{F}_1.$$

Obviously, by the condition (A3),

$$\begin{aligned}
&M_n^{-2}E||g(Y,X,u_0,\beta)||^2 \\
&\le chM_n^{-2}\sup_{u_0,t,||\beta-\beta_0||\le r_0} E_{Y|U=u_0+th}G_k^2(Y-\beta^\tau Z(X,t))||Z(X,t)||^2 \\
&\le O(hM_n^{-2}) = v.
\end{aligned}$$

Now let $M_n = n^{s_0}$, $\delta_n = (h^{1/2}\wedge n^{-1/\xi})(\log n)^{-1}$ and $M = \delta_n n^{1/2}hM_n^{-1}$. Using Lemma 6.2 in Zhang and Gijbels (1999), we have

$$\begin{aligned}
&P\{\sup||A_{n1}(u_0,\beta)||\delta_n^{-1} > M_0\} = P\{\sup_{g\in\mathcal{F}_1}||v_n(gM_n^{-1})|| > M) \\
&\le c_1(n^{1/2}(MhM_n)^{-1})^{w_1}\exp\{-c_3M^2/v\} + c_2v^{-w_1}\exp(-nv) \\
&= O((h^2\delta_n)^{-w_1})\exp\{-c_3\delta_n^2nh^2M_n^{-2}/hM_n^{-2}\} + c_2O(hM_n^{-2})^{-w_1}\exp(-c_4nhM_n^{-2}).
\end{aligned} \tag{4.2}$$

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

$$\begin{aligned}
&b_0\le hn^\eta\le b_1, \quad nh^2/\log n\to\infty, \quad n^{1-2/\xi}h/\log n\to\infty, \\
&nhM_n^{-2}/\log n\to\infty, \quad M_n^{-\xi+1}\delta_n^{-1}\to 0
\end{aligned}$$

The above requirements are fulfilled provided that for $s_0>0$,

$$\begin{aligned}
&b_0\le hn^\eta\le b_1, \quad 0<\eta<\min\{\frac{1}{2}, 1-2/\xi\}, \\
&\max\{\eta/(2(\xi-1)), 1/(\xi(\xi-1))\} < s_0 < (1-\eta)/2.
\end{aligned}$$

These conditions are equivalent to

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

since $\xi\ge 4$. 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 $n \to \infty$, $b_0 \le hn^\eta \le b_1$, $0 < \eta < 1/2$, we have*

$$\begin{aligned} V_n(u_0,\beta) &= EK_h(U-u_0)\mathbf{G}_h(u_0,\beta)\mathbf{G}_h^\tau(u_0,\beta) + o_p(h^{1/2}) \\ &= V(u_0)\otimes(S\otimes\Gamma(u_0)) + o_p(h^{1/2}) + O(\|\beta-\beta_0\|). \end{aligned}$$

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

$$\begin{aligned} &\|EK_h(U-u_0)\mathbf{G}_h(u_0,\beta)\mathbf{G}_h^\tau(u_0,\beta)I(F(Y,X) > M_n)\| \\ &\le \sup_u E[F(Y,X)^\xi | U = u] M_n^{2-\xi} = O(M_n^{2-\xi}). \end{aligned}$$

Let $M_n = n^{s_0}$. Then a similar inequality to (4.2) holds if

$$\begin{aligned} &b_0 \le hn^\eta \le b_1, \quad nh^2/\log n \to \infty, \\ &nhM_n^{-4}/\log n \to \infty, \quad M_n^{2-\xi}h^{-1/2} \to 0. \end{aligned}$$

These conditions are equivalent to

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

due to the fact that $\xi \ge 4$. 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 $n \to \infty$, $b_0 \le hn^\eta \le b_1$, $1/(2\xi) < \eta < 1/2$, then we have*

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

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

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

Note that the condition (A2) implies

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

uniformly in $u_0 \in \Omega$ and $\|\beta - \beta_0\| \le r_0$. Combining (4.4) with (4.3), Lemma 4.2 and the condition (A9), and and using the same argument as that of Owen (1988), we have

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

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

$$u(u_0,\beta)\|EK_h(U-u_0)\mathbf{G}_h(u_0,\beta)\| = EK_h(U-u_0)\mathbf{G}_h(u_0,\beta).$$

Define

$$\begin{aligned} T_n(u_0,\beta) &= \frac{1}{n}\sum_{i=1}^n K_h(U_i-u_0)\log(1+\phi_n u(u_0,\beta)^\tau \mathbf{G}_{ih}(u_0,\beta)), \\ T_{n1}(u_0,\beta) &= \frac{1}{n}\sum_{i=1}^n K_h(U_i-u_0)\log(1+\phi_n u(u_0,\beta)^\tau \mathbf{G}_{ih}(u_0,\beta))I(\|\mathbf{G}_{ih}\| \le n^{1/\xi}). \end{aligned}$$

Then uniformly for u_0 and β,

$$\begin{aligned} T_{n1}(u_0,\beta) &= \phi_n \frac{1}{n}\sum_{i=1}^{n} K_h(U_i - u_0)u(u_0,\beta)^\tau \mathbf{G}_{ih}(u_0,\beta)I(\|\mathbf{G}_{ih}\| \le n^{1/\xi}) \\ &\quad -\frac{1}{2}\phi_n^2|O(1)|\sum_{i=1}^{n} F(Y_i,X_i)^2 \\ &= \phi_n(u(u_0,\beta)^\tau EK_h(U-u_0)\mathbf{G}_h(u_0,\beta) + o_p(\phi_n^2) + O_p(\phi_n^2). \end{aligned} \tag{4.6}$$

Futhermore, by (4.3)~(4.6),

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

Consequently, we have

$$\begin{aligned} &P\{\sup_{\|\beta-\beta_0\|\ge\rho}(-T_n(u_0,\beta)) > -T_n(u_0,\beta_0), \text{for some } u_0\} \\ &\le P\{\inf_{\|\beta-\beta_0\|\ge\rho}\|EK_h(U-u_0)G_h(u_0,\beta)\| \le |O_p(\phi_n)|, \text{for some } u_0\} \end{aligned}$$

which together with the condition (A10) leads to

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

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

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

uniformly in u_0. This completes the proof.

Lemma 4.4 *Suppose for some positive constants b_0 and b_1, $b_0 \le hn^\eta \le b_1$, $0 < \eta < 1/2$. 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 $u_0 \in \Omega$, $\|\alpha\| + \|\beta - \beta_0\| \le o(n^{-1/\xi} \wedge h^{1/2})$.

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

$$\sup_{u_0\in\Omega,\|\beta-\beta_0\|\le r_0} Z_n(u_0,\beta) = o_p(n^{-1/\xi})$$

which together with Lemma 2.2 yields

$$\begin{aligned} V_n(u_0,\alpha,\beta) &= V_n(u_0,\beta) + O_p(\|\alpha\|)\frac{(2+o_p(1))}{(1+o_p(1))}\frac{1}{n}\sum_{i=1}^{n} K_h(U_i-u_0)F(Y_i,X_i)^3 \\ &= V_n(u_0,\psi_2) + O_p(\|\alpha\|) \\ &= V(u_0)\otimes(S\otimes\Gamma(u_0)) + o_p(h^{1/2}) + O_p(\|\alpha\|). \end{aligned}$$

The proof is completed.

Lemma 4.5 *Suppose there exist positive constants b_0, b_1 and η such that $b_0 \le hn^\eta \le b_1$, $0 < \eta < 1/2$. Then under the conditions (K0), (U0), (A2), (B1)~(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 $u_0 \in \Omega$, $||\alpha|| + ||\beta - \beta_0|| \le o(n^{-1/\xi} \wedge h^{1/2})$.

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

$$\max_i \sup_{u_0,\beta} |\alpha^\tau G_{ih}(u_0, \beta)| = o_p(1)$$

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

$$B_n(u_0, 0, \beta) = B_{n1}(u_0, \beta) + B_{n2}(u_0, \beta) + EK_h(U - u_0)\frac{\partial G_h(u_0, \beta)}{\partial \beta^\tau}$$

where

$$\begin{aligned} B_{n1}(u_0, \beta) &= \frac{1}{n}\sum_{i=1}^n K_h(u_i - u_0)\frac{\partial \mathbf{G}_{ih}(u_0, \beta)}{\partial \beta^\tau} I(F_4(Y, X) \le M_n) \\ &\quad - EK_h(U - u_0)\frac{\partial \mathbf{G}_h(u_0, \beta)}{\partial \beta^\tau} I(F_4(Y, X) \le M_n), \\ B_{n2}(u_0, \beta) &= \frac{1}{n}\sum_{i=1}^n K_h(u_i - u_0)\frac{\partial \mathbf{G}_{ih}(u_0, \beta)}{\partial \beta^\tau} I(F_4(Y, X) > M_n) \\ &\quad - EK_h(U - u_0)\frac{\partial \mathbf{G}_h(u_0, \beta)}{\partial \beta^\tau} I(F_4(Y, X) > M_n). \end{aligned}$$

Obviously,

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

$$E\left(K((U - u_0)/h)||\partial \mathbf{G}_h(u_0, \beta)/\partial \beta^\tau||/M_n\right)^2 \le O(hM_n^{-2}).$$

Thus, a sufficient condition for (4.7) is

$$\begin{aligned} b_0 &\le hn^\eta \le b_1, \quad M_n = n^{s_0}, \\ nh^2/\log n &\to \infty, \quad nhM_n^{-2}/\log n \to \infty, \quad M_n^{1-\nu}h^{-1/2} \to 0. \end{aligned}$$

These are equivalent to

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

by noting that $\nu \ge 2$.

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

Proof: Note that by the condition (A2) and $||\psi_1|| \leq o(n^{-1/\xi} \wedge h^{1/2})$, we have

$$\max_i \sup_{\beta, u_0} ||\alpha^\tau \mathbf{G}_{ih}(u_0, \beta)|| = o_p(1).$$

Thus

$$\begin{aligned} ||C_n(u_0, \alpha, \beta)|| &\leq O_p(||\alpha||) \frac{1}{n} \sum_{i=1}^n K_h(U_i - u_0) F_4(Y_i, X_i) F(Y_i, X_i) \\ &= O_p(||\alpha||) \end{aligned}$$

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

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

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

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

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

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n K_h(U_i - u_0) \frac{\mathbf{G}_{ih}(u_0, \hat{\beta})}{1 + \hat{\alpha}^\tau \mathbf{G}_{ih}(u_0, \hat{\beta})}, \\ 0 &= \frac{1}{n} \sum_{i=1}^n K_h(U_i - u_0) \frac{\hat{\alpha}^\tau \partial \mathbf{G}_{ih}(u_0, \hat{\beta}) / \partial \beta^\tau}{1 + \hat{\alpha}^\tau \mathbf{G}_{ih}(u_0, \hat{\beta})}. \end{aligned}$$

Then the Taylor expansion, we have

$$\begin{aligned} 0 &= A_n(u_0, \beta_0) - V_n(u_0, \alpha_{n1}, \beta_{n1})\hat{\alpha} + (B_n(u_0, \alpha_{n1}, \beta_{n1}) - C_n(u_0, \alpha_{n1}, \beta_{n1}))(\hat{\beta} - \beta_0) \\ 0 &= \{B_n(u_0, \alpha_{n2}, \beta_{n2} - C_n(u_0, \alpha_{n2}, \beta_{n2})\}\hat{\alpha} + (D_n(u_0, \alpha_{n2}, \beta_{n2}) - E_n(u_0, \alpha_{n2}, \beta_{n2})(\hat{\beta} - \beta_0) \end{aligned}$$

where α_{nj}, $j = 1, 2$ are between $\hat{\alpha}$ and 0 and β_{nj}, $j = 1, 2$, are between $\hat{\beta}$ and β_0. By using Lemmas 4.4 to 4.8, the above equations become

$$\begin{aligned} -A_n(u_0, \beta_0) &= -(1 + o_p(h^{1/2}))V(u_0) \otimes (S \otimes \Gamma(u_0))\hat{\alpha} + (o_p(h^{1/2}) + D(u_0) \otimes (S \otimes \Gamma(u_0)))(\hat{\beta} - \beta_0), \\ 0 &= (o_p(h^{1/2}) + D(u_0) \otimes (S \otimes \Gamma(u_0)))\hat{\alpha} + o_p(h^{1/2})(\hat{\beta} - \beta_0). \end{aligned}$$

It follows that

$$\begin{aligned} (\hat{\beta} - \beta_0) &= -[(D(u_0)^\tau V^{-1}(u_0)D(u_0))^{-1}D(u_0)V^{-1}(u_0) \otimes (S^{-1} \otimes \Gamma(u_0)^{-1}) + o_p(h^{1/2})]A_n(u_0, \beta_0), \\ \hat{\alpha} &= [V^{-1}(u_0) - V^{-1}(u_0)(D^\tau(u_0)V^{-1}(u_0)D(u_0))^{-1}D(u_0)D^\tau(u_0)V^{-1}(u_0) + o_p(h^{1/2})]A_n(u_0, \beta_0). \end{aligned}$$

Observe that

$$\begin{aligned} A_n(u_0, \beta) &= \frac{1}{n}\sum_{i=1}^n K_h(U_i - u_0)G(Y_i - A^\tau(U_i)X_i + \frac{1}{2}A''(U_i^*)X_i(U_i - u_0)^2 \\ &\quad +(\beta - \beta_0)^\tau Z(X_i, (U_i - u_0)/h)) \otimes Z(X_i, (U_i - u_0)/h) \\ &= \frac{1}{n}\sum_{i=1}^n K_h(U_i - u_0)G(\varepsilon_i) \otimes Z(X_i, (U_i - u_0)/h) + \frac{h^2}{2}O_p(1) + O_p(||\beta - \beta_0||) \end{aligned}$$

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, ε and X are indepedent 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,

$$\begin{aligned} V_n^*(U_j) &= V(U_j) \otimes (S \otimes \Gamma(U_j))(1 + o_p(h^{1/2})) \\ \hat{\alpha}(U_j) &= V_n^*(U_j)^{-1}\frac{1}{n}\sum_{i=1}^n K_h(U_i - U_j)\mathbf{G}_{ih}(U_j, \hat{\beta}). \end{aligned}$$

The last two equalities lead to

$$\begin{aligned} l(G) &= \sum_{j=1}^n \hat{\alpha}(U_j)^\tau \sum_{i=1}^n \frac{K_h(U_i - U_j)}{\sum_{m=1}^n K_h(U_m - U_j)}\mathbf{G}_{ih}(U_j, \hat{\beta}) \\ &\quad -\frac{1}{2}\sum_{j=1}^n \hat{\alpha}(U_j)^\tau V_n(U_j, s^*\hat{\alpha}(U_j), \hat{\beta})\hat{\alpha}(U_j) \\ &= \sum_{j=1}^n \hat{\alpha}(U_j)^\tau[\frac{1}{\sum_{m=1}^n K_h(U_m - U_j)}V_n^*(U_j) - \frac{1}{2}V_n(U_j, s^*\hat{\alpha}(U_j, \hat{\beta}))]\hat{\alpha}(U_j) \\ &= \frac{1}{2}(1 + o_p(h^{1/2}))\sum_{j=1}^n f^{-1}(U_j)\hat{\alpha}(U_j)^\tau[V(U_j) \otimes (S \otimes \Gamma(u_0))]\hat{\alpha}(U_j), \end{aligned} \tag{4.8}$$

where $0 \le s^* \le 1$, and $V_n(u, \alpha, \beta)$ is defined at Subsection 4.1. Note that we draw out the factor $1 + o_p(h^{1/2})$ from the inside of the summation in (4.8) because the $o_p(h^{1/2})$ is uniform with respect to U_j, $1 \le j \le n$ and $\hat{\alpha}(U_j)^\tau[V(U_j) \otimes (S \otimes \Gamma(U_j))]\hat{\alpha}(U_j)/\hat{\alpha}(U_j)^\tau\hat{\alpha}(U_j)$, $1 \le j \le n$ are bounded away from 0 and ∞ (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

$$
\begin{aligned}
l(G) &= (\frac{1}{2}+o_p(h^{1/2}))\sum_{j=1}^{n}\frac{1}{n}\sum_{i=1}^{n}K_h(U_i-U_j)(C(U_j)G(\varepsilon_i))^{\tau} \\
&\otimes\begin{pmatrix} \Gamma^{-1}(U_j)X_i \\ \mu_2^{-1}(U_i-U_j)\Gamma^{-1}(U_j)X_i/h \end{pmatrix}^{\tau}\frac{1}{f(U_j)}[V(U_j)\otimes(S\otimes\Gamma(U_j))] \\
&\times\frac{1}{n}\sum_{k=1}^{n}K_h(U_k-U_j)(C(U_j)G(\varepsilon_k))\otimes\begin{pmatrix} \Gamma^{-1}(U_j)X_k \\ \mu_2^{-1}(U_k-U_j)\Gamma^{-1}(U_j)X_k/h \end{pmatrix}+\zeta_n \\
&= (1+o_p(h^{1/2}))\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}K_h(U_i-U_j)K_h(U_k-U_j)f^{-1}(U_j) \\
&\quad\times G^{\tau}(\varepsilon_i)C(U_j)G(\varepsilon_k)(1+\mu_2^{-1}(U_i-U_j)(U_k-U_j)/h^2)X_i^{\tau}\Gamma^{-1}(U_j)X_k+\zeta_n \qquad (4.9)
\end{aligned}
$$

where $\zeta_n=0$ when A_0 is linear, and otherwise $\zeta_n=O_p(nh^4)$. The last term in (4.9) can be decomposed as follows:

$$
(1+o_p(h^{1/2}))L(G)=T_{n11}+T_{n121}+T_{n122}+T_{n21}+T_{n22}+\zeta_n \qquad (4.10)
$$

where

$$
\begin{aligned}
T_{n11} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}K_h(U_i-U_j)^2f^{-1}(U_j)[G^{\tau}(\varepsilon_i)C(U_j)G(\varepsilon_i)-E[G^{\tau}(\varepsilon_i)C(U_j)G(\varepsilon_i)|(U_i,U_j)]] \\
&\quad\times(1+\mu_2^{-1}(U_i-U_j)^2/h^2)X_i^{\tau}\Gamma^{-1}(U_j)X_i, \\
T_{n121} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}K_h(U_i-U_j)^2E[G^{\tau}(\varepsilon_i)C(U_j)G(\varepsilon_i)|(U_i,U_j)] \\
&\quad\times(1+\mu_2^{-1}(U_i-U_j)^2/h^2)[X_i^{\tau}\Gamma^{-1}(U_j)X_i-E[X_i^{\tau}\Gamma^{-1}(U_j)X_i|(U_i,U_j)]])f^{-1}(U_j), \\
T_{n122} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}K_h(U_i-U_j)^2E[G^{\tau}(\varepsilon_i)C(U_j)G(\varepsilon_i)|(U_i,U_j)] \\
&\quad\times(1+\mu_2^{-1}(U_i-U_j)^2/h^2)E[X_i^{\tau}\Gamma^{-1}(U_j)X_i|(U_i,U_j)]f^{-1}(U_j), \\
T_{n21} &= \frac{1}{n^2}\sum_{i\neq k}\sum_{j\notin\{i,k\}}K_h(U_i-U_j)K_h(U_k-U_j)G^{\tau}(\varepsilon_i)C(U_j)G(\varepsilon_k) \\
&\quad\times(1+(U_i-U_j)(U_k-U_j)\mu_2^{-1}/h)X_i^{\tau}\Gamma^{-1}(U_j)X_k, \\
T_{n22} &= \frac{K(0)}{n^2h^2}\sum_{i\neq k}^{n}\{K((U_k-U_i)/h)G^{\tau}(\varepsilon_i)C(U_j)G(\varepsilon_k)X_i^{\tau}\Gamma^{-1}(U_i)X_kf^{-1}(U_i) \\
&\quad+K((U_i-U_k)/h)G^{\tau}(\varepsilon_i)C(U_k)G(\varepsilon_k)X_i^{\tau}\Gamma^{-1}(U_k)X_kf^{-1}(U_k)\}.
\end{aligned}
$$

Observe that as $nh^{3/2}\to\infty$, $h\to 0$,

$$
\begin{aligned}
T_{n122} &= \frac{K(0)^2}{(nh)^2}\sum_{i=1}^{n}tr(C(U_i)V(U_i))pf^{-2}(U_i) \\
&\quad\frac{1}{n^2}\sum_{i\neq j}K_h(U_i-U_j)^2tr(C(U_j)V(U_i)) \\
&\quad\times(1+\mu_2^{-1}(U_i-U_j)^2/h^2)tr(\Gamma^{-1}(U_j)\Gamma(U_i))(f(U_i)f(U_j))^{-1} \\
&= \frac{K(0)^2}{nh^2}[Etr(C(U)V(U))/f^2(U)+O_p(n^{-1/2})]+\Psi_n \\
&= o_p(h^{-1/2})+\Psi_n \qquad (4.11)
\end{aligned}
$$

where

$$\begin{aligned}\Psi_n &= \frac{1}{n^2}\sum_{i\neq j} K_h(U_i-U_j)tr(C(U_j)V(U_i))\\ &\quad \times(1+\mu_2^{-1}(U_i-U_j)^2/h^2)tr(\Gamma^{-1}(U_j)\Gamma(U_i))(f(U_i)f(U_j))^{-1},\\ &= \frac{(k_0-1)p|\Omega|}{h}\int K^2(t)(1+\mu_2^{-1}t^2)dt+o_p(h^{-1/2}).\end{aligned}$$

This is because

$$\begin{aligned}E\Psi_n &= (1+O(h))\frac{p}{h}\int K(t)^2(1+\mu_2^{-1}t^2)dtEtr(C(U)V(U)f^{-1}(U)\\ &= \frac{p(k_0-1)}{h}(1+O(h))|\Omega|\int K^2(t)(1+\mu_2^{-1}t^2)dt,\\ Var(\Psi_n) &\leq O(n^{-1}h^{-2})=o(h^{-1}).\end{aligned}$$

By a similar argument, we have the following equalities

$$\begin{aligned}T_{n121} &= \frac{K(0)^2}{(nh)^2}\sum_{i=1}^n tr(C(U_i)V(U_i))(X_i^\tau\Gamma^{-1}(U_i)X_i-E[X_i^\tau\Gamma^{-1}(U_i)X_i|U_i])f^{-1}(U_i)\\ &\quad \frac{1}{n^2}\sum_{i\neq j}K_h(U_i-U_j)^2tr(C(U_j)V(U_j))(1+\mu_2^{-1}(U_i-U_j)^2/h^2)\\ &\quad \times(X_i^\tau\Gamma^{-1}(U_j)X_i-E[X_i^\tau\Gamma^{-1}(U_j)X_i|(U_i,U_j)]f^{-1}(U_j)\\ &= \frac{K(0)^2}{nh^2}O_p(n^{-1/2})+o_p(h^{-1/2}); \end{aligned}\tag{4.12}$$

$$T_{n22}=o_p(h^{-1/2}); \tag{4.13}$$

$$T_{n21}=o_p(h^{-1/2})+\frac{n-2}{n^2}\sum_{i\neq k}G^\tau(\varepsilon_i)\Phi_{ikh}G(\varepsilon_k) \tag{4.14}$$

where Φ_{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)~(4.14) imply (3.2). (3.3) can be proved by a similar argument by showing that

$$l(A_0|G)=(1+o_p(h^{1/2}))\frac{1}{2}\frac{1}{n^2}\sum_{j=1}^n A_n^\tau(U_j,\beta_0)[V(U_j)\otimes(S\otimes\Gamma(U_j))]^{-1}A_n(U_j,\beta_0).$$

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,

$$\begin{aligned}Var(T_n) &= \frac{(2+o(1))}{n(n-1)}E\{G^\tau(\varepsilon_1)\Phi_{12h}G(\varepsilon_2)\}^2\\ &= \frac{(2+o(1))}{n(n-1)}tr\{E(\Phi_{12h}G(\varepsilon_2)G^\tau(\varepsilon_2)\Phi_{123h}^\tau G(\varepsilon_1)G^\tau(\varepsilon_1))\}\\ &= \frac{2(1+O(h))}{n(n-1)}tr\{EK_h^*(U_2-U_1)^2C(U_1)G(\varepsilon_2)G^\tau(\varepsilon_2)C(U_1)G(\varepsilon_2)G(\varepsilon_2)^\tau\\ &\quad \times X_1^\tau\Gamma^{-1}(U_1)X_2X_2^\tau\Gamma^{-1}(U_1)X_1\}\end{aligned}$$

$$
\begin{aligned}
&= \frac{2(1+O(h))}{n(n-1)h} tr\{E\int K^*(t)^2 dt C(U_1)V(U_2)C(U_1)V(U_1)tr(\Gamma^{-1}(U_1)\Gamma(U_2))(f(U_2)f(U_1))^{-1}\} \\
&= \frac{2(1+O(h))p}{n(n-1)h} E\{f^{-1}(U_1)tr(C(U_1)V(U_1))\}\int K^*(t)^2 dt \\
&= \frac{2(1+O(h))}{n(n-1)} \frac{p(k_0-1)}{h} |\Omega| \int K^*(t)^2 dt. \qquad (4.15)
\end{aligned}
$$

Let $D_i = (\varepsilon_i, X_i, U_i)$, $1 \le i \le n$, and Π_k be the $\sigma-$algebra generated by $D_1, \cdots, D_k$, $1 \le k \le n$. Set $\Phi_h(D_i, D_k) = G^\tau(\varepsilon_i)\Phi_{ikh}G(\varepsilon_k)$, $\eta_{n1} = 0$, and

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

Then

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

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

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

and

$$Var^{-2}(T_n)\sum_{k=1}^{n} E\eta_{nk}^4 \longrightarrow 0. \qquad (4.17)$$

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 $i < j$,

$$
\begin{aligned}
E[\Phi_h(D_i, D_j)^2|D_i] &= \frac{1}{h}\int K^*(t)^2 dt X_i^\tau \Gamma^{-1}(U_i)X_i G^\tau(\varepsilon_i)C(U_i)G(\varepsilon_i)(1+O(h)), \\
E[\Phi_h(D_i, D)\Phi_h(D_j, D)|(D_i, D_j)] &= G(\varepsilon_i)^\tau E[K_h(U-U_i)K_h(U-U_k)C(U_i)V(U)C(U_j) \\
&\quad \times tr(\Gamma^{-1}(U_k)XX^\tau\Gamma^{-1}(U_j)X_iX_j^\tau)|(D_i, D_j)]G(\varepsilon_j), \\
E\Phi_h^2(D_i, D)\Phi_h^2(D_j, D) &= \frac{1}{h^2}(1+O(h))\int\int K^*(t)^2K^*(s)^2 dt ds E[(X_j^\tau\Gamma^{-1}(U_j)X_j)^2 \\
&\quad \times (G^\tau(\varepsilon_j)C(U_j)G(\varepsilon_j))^2 \\
E\Phi_h^4(D_i, D_k) &= O(1)(1+O(h))\frac{1}{h^3}\int K^*(t)^4 dt.
\end{aligned}
$$

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

$$
\begin{aligned}
\sum_{k=2}^{n} E[\eta_{nk}^2|\Pi_{k-1}] &= \sum_{k=2}^{n}\frac{4}{n^2(n-1)^2}\{\sum_{j=1}^{k-1} E[\Phi_h(D_j, D_k)^2|D_j] \\
&\quad + \sum_{i\neq j}^{k-1} E[\Phi_h(D_i, D_k)\Phi_h(D_j, D_k)|(D_i, D_j)]\} \\
&= \sum_{k=2}^{n}\frac{4}{n^2(n-1)^2}\sum_{i=1}^{k-1}\frac{1+O(h)}{h}\int K^*(t)^2 dt X_i^\tau\Gamma^{-1}(U_i)X_iG^\tau(\varepsilon_i)C(U_i)G(\varepsilon_i)
\end{aligned}
$$

$$
\begin{aligned}
& +\sum_{k=2}^{n}\frac{8}{n^2(n-1)^2}\sum_{i<j}^{n-1}(n-j)E[\Phi_h(D_i,D)\Phi_h(D_j,D)|(D_i,D_j)] \\
= & \ \frac{(1+O(h))4\int K^*(t)^2dt}{n^2(n-1)^2h}\sum_{i=1}^{n-1}(n-i)X_i^\tau\Gamma^{-1}(U_i)X_iG^\tau(\varepsilon_i)C(U_i)G(\varepsilon_i)+\Upsilon_n \\
= & \ (1+o(1))\frac{2\int K^*(t)^2dt}{n(n-1)}EE[X_1^\tau\Gamma^{-1}(U_i)X_i|U_i]E[G^\tau(\varepsilon_i)C(U_i)G(\varepsilon_i)|U_i]+\Upsilon_n \\
= & \ (1+o(1))Var(T_n)+\Upsilon_n
\end{aligned} \tag{4.18}
$$

where

$$
\begin{aligned}
\Upsilon_n = & \ \frac{8}{n^2(n-1)^2}\sum_{i<j}^{n-1}(n-j)G(\varepsilon_i)^\tau E[K_h^*(U-U_i)K_h^*(U-U_k)C(U_i)V(U)C(U_k) \\
& \times tr(\Gamma^{-1}(U_k)XX^\tau\Gamma^{-1}(U_i)X_iX_k^\tau)|(U_i,U_k,X_i,X_k)]G(\varepsilon_k).
\end{aligned}
$$

Note that

$$
\begin{aligned}
E[\Upsilon_n]^2 = & \ \frac{64}{n^4(n-1)^4}\sum_{i<k}^{n-1}(n-k)^2E[G(\varepsilon_i)^\tau E[K_h^*(U-U_i)K_h^*(U-U_k)C(U_i)V(U)C(U_k) \\
& \times tr(\Gamma^{-1}(U_k)XX^\tau\Gamma^{-1}(U_i)X_iX_k^\tau)|(U_i,X_i,U_k,X_k)]G(\varepsilon_k)]^2 \\
= & \ O(\frac{(k_0-1)p}{(n-1)^4h})\int[K^*(t)*K^*(t)]^2dt \\
= & \ O(\frac{1}{n^2})Var(T_n)),
\end{aligned}
$$

which implies $\Upsilon_n = o_p(Var(T_n)$ 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

$$
\begin{aligned}
\sum_{k=2}^{n}E\eta_{nk}^4 = & \ \frac{O(1)}{n^4(n-1)^4}\sum_{k=2}^{n}\sum_{i\neq j}^{k-1}\{O(\frac{1}{h^2})+O(\frac{(k-1)}{h^3})\} \\
= & \ \frac{O(1)}{n^4(n-1)^4}\{2\sum_{i<j}^{n-1}(n-j)O(\frac{1}{h^2})+O(\frac{1}{h^3})\frac{n^2}{2}\} \\
= & \ O(Var(T_n)^2)\frac{1}{(n-1)^2}\{O(n)+O(\frac{1}{h})\}.
\end{aligned}
$$

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 $\Gamma(u_0) = E[XX^\tau|U=u_0]f(u_0)$ 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:

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

$$\hat{\alpha}^*(u_0) = \frac{1}{n}\sum_{i=1}^{n} K_h(U_i-u_0)\{(V^{-1}(u_0)G(\varepsilon_i))\otimes\begin{pmatrix}\Gamma^{-1}(u_0)X_i\\ \mu_2^{-1}\Gamma^{-1}(u_0)X_i(U_i-u_0)/h\end{pmatrix}$$
$$-V^{-1}D(D^\tau V^{-1}D)^{-1}D^\tau V^{-1}G(\varepsilon_i)\otimes\begin{pmatrix}0\\ \Gamma_{22}^{-1}(u_0)X_i^{(2)}\\ 0\\ \mu_2^{-1}\Gamma_{22}^{-1}(u_0)X_i^{(2)}(U_i-u_0)/h\end{pmatrix}\}(1+o_p(h^{1/2}))+O_p(h^2).$$

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

$$\begin{aligned}X_i^\tau\Gamma(U_k)^{-1}X_k &= \{X_i^{(1)\tau}-X_i^{(2)\tau}\Gamma_{22}(U_k)\Gamma_{21}(U_k)\}\Gamma_{11,2}^{-1}(U_k)(X_k^{(1)}-\Gamma_{12}(U_k)\Gamma_{22}^{-1}(U_k)X_k^{(2)})\\ &\quad +X_i^{(2)\tau}\Gamma_{22}(U_k)^{-1}X_k^{(2)}.\end{aligned}$$

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

$$\begin{aligned}l(H_{0s}|G) &= (1+o_p(h^{1/2}))\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}K_h(U_i-U_j)K_h(U_k-U_j)\frac{1}{f(U_j)}G^\tau(\varepsilon_i+A(U_i)^\tau X_i)\\ &\quad\times V^{-1}(U_j)G(\varepsilon_k+A(U_k)^\tau X_k)(1+\mu_2^{-1}\frac{U_i-U_j}{h}\frac{U_k-U_j}{h})X_i^\tau\Gamma^{-1}(U_j)X_k-l_G\\ &= (1+o_p(h^{1/2}))(W_{n0}+W_{n1}+W_{n2}+W_{n3})-l_G \qquad (4.19)\end{aligned}$$

where

$$\begin{aligned}W_{n0} &= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}K_h(U_i-U_j)K_h(U_k-U_j)(1+\mu_2^{-1}\frac{U_i-U_j}{h}\frac{U_k-U_j}{h})\frac{1}{f(U_j)}\\ &\quad\times G(\varepsilon_i)^\tau V^{-1}(U_j)G(\varepsilon_k)X_i^\tau\Gamma^{-1}(U_j)X_k,\\ W_{n1} &= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}K_h(U_i-U_j)K_h(U_k-U_j)(1+\mu_2^{-1}\frac{U_i-U_j}{h}\frac{U_k-U_j}{h})\frac{1}{f(U_j)}\\ &\quad\times G(\varepsilon_i)^\tau V^{-1}(U_j)\frac{\partial G(\varepsilon_k^*)}{\partial\varepsilon}X_i^\tau\Gamma^{-1}(U_j)X_kX_k^\tau A(U_k),\\ W_{n2} &= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}K_h(U_i-U_j)K_h(U_k-U_j)(1+\mu_2^{-1}\frac{U_i-U_j}{h}\frac{U_k-U_j}{h})\frac{1}{f(U_j)}\\ &\quad\times\frac{\partial G(\varepsilon_i^*)^\tau}{\partial\varepsilon}V^{-1}(U_j)G(\varepsilon_k)X_i^\tau\Gamma^{-1}(U_j)X_kX_k^\tau A(U_k),\\ W_{n3} &= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}K_h(U_i-U_j)K_h(U_k-U_j)(1+\mu_2^{-1}\frac{U_i-U_j}{h}\frac{U_k-U_j}{h})\frac{1}{f(U_j)}\\ &\quad\times\frac{\partial G(\varepsilon_i^*)^\tau}{\partial\varepsilon}V^{-1}(U_j)\frac{\partial G(\varepsilon_k^*)}{\partial\varepsilon}A(U_i)^\tau X_iX_i^\tau\Gamma^{-1}(U_j)X_kX_k^\tau A(U_k)\end{aligned}$$

where ε_i^* is between ε_i and $\varepsilon_i+A(U_i)^\tau X_i$ and ε_k^* is between ε_k and $\varepsilon_k+A(U_k)^\tau X_k$. Under some reguarlity conditions,

$$W_{n1} = \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}G(\varepsilon_i)^\tau\{\sum_{j=1}^{n}K_h(U_i-U_j)K_h(U_k-U_j)(1+\mu_2^{-1}\frac{U_i-U_j}{h}\frac{U_k-U_j}{h})\frac{1}{f(U_j)}V^{-1}(U_j)$$

$$
\begin{aligned}
&\times X_i^\tau \Gamma^{-1}(U_j) X_k X_k^\tau A(U_k)\} \frac{\partial G(\varepsilon_k^*)}{\partial \varepsilon} \\
&= \frac{1}{2} W_{1n}^* + o_p(h^{-1/2}), \\
W_{n2} &= W_{n1}
\end{aligned}
$$

where W_{1n}^* is defined in (3.4). Similarly we write

$$W_{n3} = W_{n31} + 2W_{n32} + W_{n33}$$

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

$$
\begin{aligned}
W_{n31} &= \frac{1}{2n^2}\sum_{i=1}^n\sum_{k=1}^n \Xi_i^\tau \sum_{j=1}^n K_h(U_i-U_j)K_h(U_k-U_j)(1+\mu_2^{-1}\frac{U_i-U_j}{h}\frac{U_k-U_j}{h})\frac{1}{f(U_j)}V^{-1}(U_j) \\
&\quad \times \Xi_k A(U_i)^\tau X_i X_i^\tau \Gamma^{-1}(U_j) X_k X_k^\tau A(U_k) \\
&= \frac{1}{2n}\sum_{i=1}^n\sum_{k=1}^n \Xi_i^\tau K_h^*(U_i-U_k)V^{-1}(U_i)\Xi_i A(U_i)^\tau X_i X_i^\tau \Gamma^{-1}(U_j) X_k X_k^\tau A(U_k) + o_p(h^{-1/2}) \\
&= O(\frac{1}{nh^2}) + W_{2n}^*/2 + o_p(h^{-1/2}), \\
W_{n32} &= \frac{1}{2n^2}\sum_{i=1}^n\sum_{k=1}^n \Xi_i^\tau \sum_{j=1}^n K_h(U_i-U_j)K_h(U_k-U_j)(1+\mu_2^{-1}\frac{U_i-U_j}{h}\frac{U_k-U_j}{h})\frac{1}{f(U_j)}V^{-1}(U_j) \\
&\quad \times E[\frac{\partial G(\varepsilon_k^*)}{\partial \varepsilon}|U_k] A(U_i)^\tau X_i X_i^\tau \Gamma^{-1}(U_j) X_k X_k^\tau A(U_k) \\
&= W_{3n}^*/2 + o_p(h^{-1/2}), \\
W_{n33} &= \frac{1}{2n^2}\sum_{i=1}^n\sum_{k=1}^n E[\frac{\partial G(\varepsilon_i^*)}{\partial \varepsilon}|U_i] \sum_{j=1}^n K_h(U_i-U_j)K_h(U_k-U_j)(1+\mu_2^{-1}\frac{U_i-U_j}{h}\frac{U_k-U_j}{h}) \\
&\quad \times \frac{1}{f(U_j)}V^{-1}(U_j)E[\frac{\partial G(\varepsilon_k^*)}{\partial \varepsilon}|U_k] A(U_i)^\tau X_i X_i^\tau \Gamma^{-1}(U_j) X_k X_k^\tau A(U_k) \\
&= O_p(\frac{1}{nh^2}) + \frac{n}{2}E\{E[\frac{\partial G(\varepsilon)}{\partial \varepsilon}|U]^\tau V^{-1}(U)E[\frac{\partial G(\varepsilon)}{\partial \varepsilon}|U]A(U)^\tau XX^\tau A(U)(1+o(1))
\end{aligned}
$$

Recall that W_{2n}^* and W_{3n}^* are in (3.5) and (3.6) respectively. When $EA(U)^\tau XX^\tau A(U) = O(\frac{1}{nh})$, we have

$$
\begin{aligned}
W_{n31} &= \frac{1}{2n}\sum_{i=1}^n\sum_{k=1}^n \Xi_i^\tau K_h^*(U_i-U_k)V^{-1}(U_i)\Xi_i A(U_i)^\tau X_i X_i^\tau \Gamma^{-1}(U_j) X_k X_k^\tau A(U_k) + o_p(h^{-1/2}) \\
&= O(\frac{1}{nh^2}) + W_{2n}^*/2 + o_p(h^{-1/2}), \\
W_{n32} &= W_{3n}^*/2 + o_p(h^{-1/2}), \\
W_{n33} &= O_p(\frac{1}{nh^2}) + \frac{n}{2}E\{E[\frac{\partial G(\varepsilon)}{\partial \varepsilon}|U]^\tau V^{-1}(U)E[\frac{\partial G(\varepsilon)}{\partial \varepsilon}|U]A(U)^\tau XX^\tau A(U)(1+o(1)).
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
l_G &= (1+o_p(h^{-1/2}))[\frac{1}{2n^2}\sum_{i=1}^n\sum_{k=1}^n\sum_{j=1}^n K_h(U_i-U_j)K_h(U_k-U_j)(1+\mu_2^{-1}\frac{U_i-U_j}{h}\frac{U_k-U_j}{h}) \\
&\quad \frac{1}{f(U_j)}G^\tau(\varepsilon_i)C(U_j)G(\varepsilon_k) + 2S_{n1} + S_{n2}] \qquad (4.20)
\end{aligned}
$$

where

$$
S_{n1} = \frac{1}{2n^2}\sum_{i=1}^n\sum_{k=1}^n\sum_{j=1}^n K_h(U_i-U_j)K_h(U_k-U_j)(1+\mu_2^{-1}\frac{U_i-U_j}{h}\frac{U_k-U_j}{h})\frac{1}{f(U_j)}
$$

$$
\begin{aligned}
& \times G^{\tau}(\varepsilon_i)C(U_j)\frac{\partial G(\varepsilon_k^*)}{\partial \varepsilon}A''(U_j^*)^{\tau}(U_k-U_j)^2X_k, \\
= \; & O_p(n(nh)^{-1}h^2) = O_p(h), \\
S_{n2} \; = \; & \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{j=1}^{n}K_h(U_i-U_j)K_h(U_k-U_j)(1+\mu_2^{-1}\frac{U_i-U_j}{h}\frac{U_k-U_j}{h})\frac{1}{f(U_j)} \\
& \times\frac{\partial G(\varepsilon_i^*)^{\tau}}{\partial \varepsilon}C(U_j)\frac{\partial G(\varepsilon_k^*)}{\partial \varepsilon} \\
& \times A''(U_j)X_iX_i^{\tau}\Gamma^{-1}(U_j)X_kX_k^{\tau}A''(U_j)(U_i-U_j)^2(U_k-U_j)^2 \\
= \; & \frac{nh^4}{8}E\{D^{\tau}(U)C(U)D(U)A''(U)^{\tau}XX^{\tau}A''(U)\} \\
& \times\int\int t^2(s+t)^2K(t)K(s+t)(1+\mu_2^{-1}t(s+t))dtds(1+o_p(1))
\end{aligned}
$$

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.sta.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

Azzalini, A. & Bowman, A.N. (1993): "On the use of nonparametric regression for checking linear relationships," *J. Roy. Statist. Soc. Ser.B*, **55**, 549-557.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J. (1993): *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore and London: The Johns Hopkins University Press.

Bickel, P.J. & Ritov, Y. (1992): "Testing for goodness of fit: a new approach," in *Nonparametric Statistics and Related Topics*, Ed. A.K.Md.E. Saleh, North-Holland, New York, pp. 51-7.

Brumback, B. and Rice, J. A. (1998): "Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion)," *J. Amer. Statist. Assoc.*, **93**, 961-994.

Cai, Z., Fan, J. and Li, R. (2000): " Efficient estimation and inferences for varying-coefficient models," *Jour. Ameri. Statist. Assoc.*, **95**, 888-902.

Cai, Z., Fan, J. and Yao, Q. (2000): "Functional-coefficient regression models for nonlinear time series," *Journal of American Statistical Association*, **95**, 941-956.

Carroll, R.J., Ruppert, D. and Welsh, A.H. (1998): "Nonparametric estimation via local estimating equations," *Jour. Ameri. Statist. Assoc*, **93**, 214-227.

Cleveland, W.S. and Grosse, E. and Shyu, W.M.(1991): "Local regression models," in *Statistical Models in S* (Chambers, J.M. and Hastie, T.J., edsi.) Wadsworth & Brooks, Pacific Grove: Wadsworth & Brooks, pp. 309-376.

Chen, R. and Tsay, R.J. (1993): "Functional-coefficient autoregressive models," *J. Amer. Statist. Assoc.*, **88**, 298–308.

Eubank, R.L. and Hart, J.D. (1992): "Testing goodness-of-fit in regression via order selection criteria," *Ann. Statist.*, **20**, 1412-1425.

Eubank, R.L. and LaRiccia, V.M. (1992): "Asymptotic comparison of Cramér-von Mises and nonparametric function estimation techniques for testing goodness-of-fit," *Ann. Statist.*, **20**, 2071-86.

Fan, J. (1996): "Test of significance based on wavelet thresholding and Neyman's truncation," *J. Amer. Statist. Assoc.*, **91**, 674–688.

Fan, J., Zhang, C. and Zhang, J. (2001): "Generalized likelihood ratio statistics and Wilks phenomenon," *Ann. Statist.*, to appear.

Fan, J. and Zhang, W. (1999): "Statistical estimation in varying coefficient models," *Ann. Statist.*, **27**, 1491-1518.

Farrell, R.H. (1985): *Multivariate Calculation,* New York: Springer.

Härdle, W. and Mammen, E. (1993): "Comparing nonparametric versus parametric regression fits," *Ann. Statist.*, **21**, 1926–47.

Hart, J.D. (1997): *Nonparametric Smoothing and Lack-of-Fit Tests.* New York: Springer-Verlag.

Hastie, T.J. and Tibshirani, R.J. (1993): "Varying-coefficient models (with discussion)," *Journal of the Royal Statistical Society, B,* **55,** 757-796.

Hettmansperger, T.P. (1984): *Statistical Inference Based on Ranks.* New York: Wiley.

Hong, Y. and Lee, T.-H. (1999): "Inference and forecast of exchange rates via generalized spectrum and nonlinear time series models," *Manuscript.*

Inglot, T. and Ledwina, T. (1996): "Asymptotic optimality of data-driven Neyman's tests for uniformity," *Ann. Statist.*, **24**, 1982–2019.

Ingster, Yu. I. (1993): "Asymptotically minimax hypothesis testing for nonparametric alternatives I–III." *Math. Methods Statist.*, **2** 85-114; **3** 171-189; **4** 249-268.

Kallenberg, W.C.M. and Ledwina, T. (1997): "Data-Driven smooth tests when the hypothesis is composite," *Jour. Ameri. Statist. Assoc.*, **92,** 1094 –1104.

Kitamura, Y. (1997): "Empirical likelihood methods with weakly dependent processes," *Ann. Statist.*, **25,** 2084-2102.

Liu, A. and Zhang, J. (2000): "Local polynomial fitting based on empirical likelihood," *Report 2000-25,* EURANDOM, the Netherlands.

Newey, W.K. (1993): "Efficient estimation of models with conditional moment restrictions," in *Handbook of Statistics,* Vol. **11**, Maddala, G.S., C.R. Rao and H.D. Vinod, eds. Amsterdam: Elsevier Science Publishers B.V., pp. 419-453.

Owen, A. B. (1988): "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika,* **75**, 237–249.

Pollard, D. (1984): *Convergence of Stochatic Processes.* New York: Springer-Verlag.

Shiryayev, A.N. (1996): *Probability Theory,* Second Edition, New York: Springer.

Spokoiny, V.G. (1996): "Adaptive hypothesis testing using wavelets," *Ann. Statist.*, **24**, 2477-2498.

Tong, H. (1990): *Non-Linear Time Series: A Dynamical System Approach.* Oxford: Oxford University Press.

Wilks, S.S. (1938): "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *Ann. Math. Stat.*, **9**, 60-62.

Wu, C.O., Chiang, C.T. and Hoover, D.R. (1998): "Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data," *J. Amer. Statist. Assoc.*, **93**, 1388-1401.

Zhang, J. and Gijbels, I. (1999): "Sieve empirical likelihood and extensions of generalized least squares," *Discussion Paper 9911,* Institute of Statistics, Universite Catholique de Louvain.