

# Generalized Likelihood Ratio Tests for Varying-Coefficient Models with Censored Data

Rong Jiang, Wei-Min Qian

Department of Mathematics, Tongji University, Shanghai, 200092, P. R. China

E-mail: jrtrying@126.com

Received February 25, 2011; revised March 15, 2011; accepted March 23, 2011

## Abstract

In this paper, we extend the generalized likelihood ratio test to the varying-coefficient models with censored data. We investigate the asymptotic behavior of the proposed test and demonstrate that its limiting null distribution follows a  $\chi^2$  distribution, with the scale constant and the number of degree of freedom being independent of nuisance parameters or functions, which is called the Wilks phenomenon. Both simulated and real data examples are given to illustrate the performance of the testing approach.

**Keywords:** Varying Coefficient Model, Generalized Likelihood Ratio Test, Local Linear Method, Wilks Phenomenon, Censoring.

## 1. Introduction

Nonparametric regression model has become one of the main approaches in modern statistics due to its robustness and wide applications. In particular, it can be well estimated when the covariate is one dimension. However, as the dimension of the covariates, we face the phenomenon called "the curse of dimensionality". The varying coefficient model, which is the function approximation method for high dimension, is proposed (see Hastie and Tibshirani, 1993). Recently many statisticians (see Fan and Zhang 1999, 2000; Cai, 2007; Zhou and Ling, 2009; Wang and Xia, 2009; Chen and Tong, 2010) have investigated the varying coefficient model due to its simplifying structure, meaningful interpretation and wide application.

The varying-coefficient model has the following form:

$$Y = \sum_{i=1}^p \alpha_i(U) X_i + \varepsilon = \boldsymbol{\alpha}(U)^T \mathbf{X} + \varepsilon \quad (1.1)$$

where  $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \dots, \alpha_p(\cdot))^T$ ,  $\alpha_1(u), \dots, \alpha_p(u)$  is unspecified smoothing function that needs to estimate,  $\mathbf{X} = (X_1, \dots, X_p)^T$  is random vector,  $U$  is random variable, and its density function is  $f(u)$ ,  $\varepsilon$  is random error, and  $E(\varepsilon|U, \mathbf{X}) = 0$ ,  $\text{Var}(\varepsilon|U, \mathbf{X}) = \sigma^2(U, \mathbf{X})$ .

However, in the real problems, for example, in the fields of reliable lifespan experiment, medicine track,

survival analysis and so on,  $Y$  can not be observed because it is censored. Let  $C$  denotes the censoring random variable,  $Y$  and  $C$  are independent random variable under the condition that  $U$  and  $\mathbf{X}$  are given.  $T = \min(Y, C)$ ,  $T_i = \min(y_i, c_i)$ ,  $\delta_i = I(y_i \leq c_i)$  ( $i = 1, 2, \dots, n$ ), where  $I(\cdot)$  denotes the sign function of an event, if  $Y$  is not censored, then  $\delta = 1$ , if  $Y$  is censored, then  $\delta = 0$ . We can only observe  $\{(U_i, X_i, T_i, \delta_i); i = 1, \dots, n\}$ . Because responding variable is censored, we can not use the methods directly which we use on full data. So we should transform the data in an unbiased way to account for the censoring. An example of this kind is given in Buckley and James (1979). However, their transformation involves the unknown regression function leading to an iterative scheme. Motivated by the Buckley-James transformation, Koul, Susarla and Van Ryzin (1981) consider a transformation which only depends on the censoring distribution, but not on the regression function. Zheng (1987) proposed a class of transformations of this type. Once such a transformation is carried through, one can apply a variety of statistical techniques to analyze the transformed data as if they were uncensored. However, since such a transformation does not involve the distribution of the response variable, it increases the variability. Hence, some smoothing technique is necessary for modeling the transformed data. Some nonparametric regression techniques were applied in

Dabrowska (1987) and Zheng (1988). In this paper, Class-K method is used to transform data.

In an effort to derive a generally applicable testing approach, Fan et al (2001) proposed the generalized likelihood ratio (GLR) statistic for nonparametric models. Their motivation was as follows. The maximum likelihood ratio test statistic in general may not exist in nonparametric and semiparametric settings. Even if it does, it is hard to find and may not be optimal in the simplest nonparametric regression setting. These drawbacks can be avoided when the maximum likelihood estimator is replaced by other reasonable nonparametric estimators, resulting in a class of statistics called the GLR statistic. The GLR test is intuitively appealing. Fan et al (2001) showed that for a variety of models and a number of nonparametric versus nonparametric and parametric versus nonparametric testing problems, the null distribution of the GLR test statistic follows an asymptotically  $\chi^2$  distribution, independent of nuisance parameters. This property is called the Wilks phenomenon and facilitates the application of the GLR statistic. The critical value can be determined either by asymptotic distributions or by simulations. In this paper, we extend the generalized likelihood ratio test to the varying-coefficient models with censored data.

The paper is organized as follows. Generalized likelihood ratio test is presented in section 2. In section 3, we provide two numerical results. Technical proofs are relegated to the Appendix.

## 2. Generalized Likelihood Ratio Tests

First, we replace the data point  $(U, X, T, \delta)$  with the transformed data point  $(U, X, Y^*)$  according to

$$Y^* = \delta\phi_1(U, X, T) + (1 - \delta)\phi_2(U, X, T) \quad (2.1)$$

where  $\phi_1(\cdot, \cdot, \cdot)$  and  $\phi_2(\cdot, \cdot, \cdot)$  are the transformation functions. In the sequel of this paper we will refer to this transformation as the "ideal transformation", since it assumes that the transformation function  $\phi_1(\cdot, \cdot, \cdot)$  and  $\phi_2(\cdot, \cdot, \cdot)$  are known. In practical situations however, those transformation functions typically have to be estimated. The estimations of  $\phi_1(\cdot, \cdot, \cdot)$  and  $\phi_2(\cdot, \cdot, \cdot)$  can be expressed as (Fan and Gijbels, 1994)

$$\begin{cases} \phi_1(u, x, y) = (1+a) \int_0^y \bar{G}^{-1}(t|u, x) dt - ay\bar{G}^{-1}(y|u, x), \\ \phi_2(u, x, y) = (1+a) \int_0^y \bar{G}^{-1}(t|u, x) dt, \end{cases} \quad (2.2)$$

where  $\bar{G}(\cdot|u, x)$  are respectively the conditional survival function of the random variable  $C$  given  $U = u$  and  $X = x$ . Remark that the Koul, Susarla and Van Ryzin (1981) transformation corresponds to  $a = -1$

and Leurgans (1987) transformation relates to  $a = 0$ . The tuning parameter  $a$  in this New Class of transformations (2.2) creates the opportunity to improvement.

Starting from the transformed data  $\{(U_i, X_i, Y_i^*) : i = 1, \dots, n\}$ , we now estimate the true regression functions  $\{\alpha_i(\cdot), i = 1, \dots, p\}$ . Fix a point  $u$ , approximate the unknown function:

$$\alpha_i(z) \approx \alpha_i(u) + \alpha_i'(u)(z - u) \equiv a_i + b_i(z - u).$$

This leads to the following weighted local least-squares problem: find  $\{(a_i, b_i), i = 1, \dots, p\}$  so as to minimize

$$\sum_{k=1}^n \left[ Y_k^* - \sum_{i=1}^p \{a_i + b_i(U_k - u)\} X_{ki} \right] K_{\hat{h}_k(u)}(U_k - u) \quad (2.3)$$

where  $K_{\hat{h}_k(u)}(\cdot) = K(\cdot / \hat{h}_k(u))$  with  $K$  being a symmetric probability density function and we use the adaptive variable bandwidth of order  $k$ , and  $\hat{h}_k(u) = (U_{l+k} - U_{l-k})/2$ , here  $l$  is the index of the design point  $U_l$  closest to  $u$ , the smoothing parameter  $k$  can be obtained by cross-validation.

Let us work with the matrix notation. Denote  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ ,  $\mathbf{X} = (X_1, \dots, X_n)^T$ ,  $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^T$ ,  $\mathbf{Z}$  denotes an  $n \times 2p$  matrix with  $(X_i^T, (U_i - u)X_i^T)$  as its  $i$ th row, and

$$\mathbf{W} = \text{diag}(K_h(U_1 - u), \dots, K_h(U_n - u))$$

The solution to the problem (2.3) is given by

$$\hat{a}_j(u) = \mathbf{e}_{j,2p}^T \{ \mathbf{Z}^T \mathbf{W} \mathbf{Z} \}^{-1} \mathbf{Z}^T \mathbf{W} \mathbf{Y}^*,$$

where  $\mathbf{e}_{j,2p}$  is the  $2p \times 1$  unit vector with 1 at the  $(j+1)$ th position.

Consider the varying-coefficient model defined in (1.1). A nature question arises in practice is if these coefficient functions are really varying. This amounts to testing the following problem:

$$H_0 : \alpha_1(U) = \theta_1, \dots, \alpha_p(U) = \theta_p,$$

where  $\theta_1, \dots, \theta_p$  are unknown parameters. Following the same derivations as in Fan et al (2001), generalized likelihood ratio tests based on local linear fits are given by

$$\lambda_n = \frac{n}{2} \log \frac{RSS_0}{RSS_1} \approx \frac{n}{2} \frac{RSS_0 - RSS_1}{RSS_1},$$

where  $RSS_1 = \sum_{k=1}^n (Y_k^* - \hat{\alpha}(U_k)^T X_k)^2$  and  $RSS_0 = \sum_{k=1}^n (Y_k^* - \hat{\theta}^T X_k)^2$  with  $\hat{\theta}$  is the least-square estimate

under the null hypothesis.

We now describe our generalized Wilks type of theorem as follows:

**Theorem 1** Suppose that conditions (C1)-(C5) given in the Appendix hold. Then, under  $H_0$ , as  $h \rightarrow 0$ ,  $nh^{3/2} \rightarrow \infty$ ,

$$r'_k \lambda_n \xrightarrow{L} \chi_{a'_n}^2.$$

where  $\xrightarrow{L}$  stands for convergence in distribution, and

$$r'_k = r_k \left[ E \sigma^{*2}(X, U) \right] \int_{\Omega} \text{tr} \left( \Gamma^*(u) \Gamma(u)^{-1} \right) du$$

$$\left[ \int_{\Omega} \text{tr} \left( \Gamma^*(u) \Gamma(u)^{-1} \right)^2 du \right]^{-1},$$

$$a'_n = r_k c_k h^{-1} \left[ \int_{\Omega} \text{tr} \left( \Gamma^*(u) \Gamma(u)^{-1} \right) du \right]^2$$

$$\left[ \int_{\Omega} \text{tr} \left( \Gamma^*(u) \Gamma(u)^{-1} \right)^2 du \right]^{-1},$$

where  $\sigma^{*2}(u, x) = \text{var}(Y^* | U = u, X = x)$ ,

$$\Gamma(u) = E(\mathbf{X}\mathbf{X}^T | U = u) f(u),$$

$$\Gamma^*(u) = E(\mathbf{X}\mathbf{X}^T \sigma^{*2}(X, U) | U = u) f(u),$$

$$c_k = K(0) - \frac{1}{2} \|K\|_2^2,$$

$$r_k = \left( K(0) - \frac{1}{2} \int K^2(t) dt \right) / \left( \int \left( K(t) - \frac{1}{2} K * K(t) \right)^2 dt \right)$$

with  $K * K$  denote the convolution of  $K$ .

### 3. Numerical Studies

In this section, we first use Monte Carlo simulation studies to assess the finite sample performance of the test procedure proposed and then demonstrate the application of the method proposed by using a real data example. The programs are written in Matlab and are available upon request from the authors.

#### 3.1. Simulation Example

Simulation data are generated from the varying-coefficient partially linear model with censored data:

$$Y = \sin(6\pi U) X_1 + \sin(2\pi U) X_2 + \beta_1 Z_1 + \varepsilon \quad (3.1)$$

where the covariate  $U$  is uniformly distributed on  $[0, 1]$ , the covariates  $X_1, X_2, Z_1$  are normally distributed with mean 0 and variance 1 and the distribution function of  $\varepsilon$  is  $F$ , where  $F = N(0, 1)$  or  $U(0, 1)$ .  $U, X_1, X_2, Z_1$  and  $\varepsilon$  are simulated independently. The censoring variable  $C \sim U(0, v)$ . We vary  $v$  to produce difference censoring rates(CR). Here 20% and

40% censoring are considered. The true parameter for  $\beta_1$  is always fixed at  $\beta_1 = 2$ , and we take  $h = n^{-1/5}$  for smooth parameter.

For this example, we draw 1000 random samples of size 100 from the model (3.1) and take  $\beta_1$  as  $\alpha_3(U)$ . We consider three null hypothesis

$$H_0^1: \alpha_1(U) = b_1; \quad H_0^2: \alpha_2(U) = b_2; \quad H_0^3: \alpha_3(U) = b_3.$$

**Table 1** and **Table 2** show that  $\alpha_1(U)$  and  $\alpha_2(U)$  are nonparametric,  $\alpha_3(U)$  is certain parametric under two different error distributions and censoring rates. The results show that the GLR test performs well.

#### 3.2. A real Data Example

We now illustrate the proposed method by an application to the chronic granulomatous disease (CGD) data set. The CGD study in a report by the International CGD Cooperative Study Group (1991), was designed to have a single interim analysis when the follow-up data as of July 15, 1989 were complete. The monitoring committee for the trial terminated the trial at a meeting on September 22, 1989. The treatment given each patient was unblinded at the first scheduled visit for the patient following the decision of the monitoring committee.

The variables contained here are:  $Z_1$ : Treatment Code, 1 = rIFN, 2 = placebo;  $Z_2$ : Pattern of inheritance, 1 = X-linked, 2 = autosomal recessive;  $Z_3$ : Age, in years;  $Z_4$ : Height, in cm;  $Z_5$ : Weight, in kg;  $Z_6$ : Using corticosteroids at time of study entry, 1 = yes, 2=no;  $Z_7$ : Using prophylactic antibiotics at time of study entry, 1 = yes, 2 = no;  $Z_8$ : 1 = male, 2 = female;  $Z_9$ : Hospital category, 1 = US-NIH, 2=US-other, 3 = Europe-Amsterdam, 4 = Europe-other;  $T_1$ : Elapsed time (in days) from randomization to diagnosis of a serious infection, or if a censored observation, elapsed time from randomization to censoring date;  $\delta$ : Censoring indicator, 1 = Non-censored observation, 2 = censored observation;  $S$ : Sequence number. For each patient, the infection records are in sequence number order.

We take  $Z_0 = 1$  as the intercept term and  $U = S / 20$ , and employ the varying-coefficient model with censored data,

$$Y = \alpha_0(U) + \sum_{i=1}^9 \alpha_i(U) Z_i + \varepsilon.$$

to fit the given data. A natural question is whether the coefficients functions are constant. To answer this question, the proposed GLR test is employed. The p-values for the test is summarized in **Table 3**. It can be seen from Table 3 that we should use following model to fit the given data.

$$Y = b_0 + \sum_{i=2}^9 b_i Z_i + \alpha_1(U) Z_1 + \varepsilon \quad (3.2)$$

**Table 1. Power of the GLR test for  $\alpha(\cdot)$  under different error distributions (DEDs) for the CR = 20%.**

$\alpha$	DEDs	$\alpha_0$	$\alpha_1$	$\alpha_2$
0.10	N(0,1)	0.542	0.996	0.000
	U(0,1)	0.646	0.999	0.000
0.05	N(0,1)	0.444	0.992	0.000
	U(0,1)	0.530	0.999	0.000
0.01	N(0,1)	0.233	0.957	0.000
	U(0,1)	0.313	0.987	0.000

**Table 2. Power of the GLR test for  $\alpha(\cdot)$  under different error distributions (DEDs) for the CR = 40%.**

$\alpha$	DEDs	$\alpha_0$	$\alpha_1$	$\alpha_2$
0.10	N(0,1)	0.509	0.970	0.000
	U(0,1)	0.534	0.991	0.000
0.05	N(0,1)	0.375	0.958	0.000
	U(0,1)	0.413	0.982	0.000
0.01	N(0,1)	0.186	0.888	0.000
	U(0,1)	0.232	0.936	0.000

**Table 3. p-values for testing whether a coefficient functions is constant.**

	$\alpha_0(U)$	$\alpha_1(U)$	$\alpha_2(U)$	$\alpha_3(U)$	$\alpha_4(U)$	$\alpha_5(U)$	$\alpha_6(U)$	$\alpha_7(U)$	$\alpha_8(U)$	$\alpha_9(U)$
GLR	0.0059	4.2791	0.0590	0.1363	0.2994	0.0242	0.3334	0.1503	0.3106	0.3600
p-values	0.9320	0.0365	0.7954	0.6975	0.5692	0.8661	0.5486	0.6836	0.5622	0.5335

And by the method of Fan and Huang (2005), we can estimate the parameters in the model (3.2).

**5. References**

[1] J. Buckley and I. R. James, "Linear Regression with Censored Data," *Biometrika*, Vol. 66, No. 3, 1979, pp. 429-436.

[2] Z. Cai, "Trending Time-Varying Coefficient Time Series Models with Serially Correlated Errors," *Journal of Econometrics*, Vol. 136, No. 1, 2007, pp. 163-188.

[3] K. Chen and X. W. Tong, "Varying Coefficient Transformation Models with Censored Data," *Biometrika*, Vol. 97, No. 4, 2010, pp. 969-976.

[4] D. M. Dabrowska, "Non-parametric Regression with Censored Survival Time Data," *Scandinavian Journal of Statistics*, Vol. 14, No. 3, 1987, pp. 181-197.

[5] J. Fan and I. Gijbels, "Censored Regression: Local Linear Approximations and Their Applications," *Journal of the American Statistical Association*, Vol. 89, No. 426, 1994, pp. 560-570.

[6] J. Fan and T. Huang, "Profile Likelihood Inferences on Semiparametric Varying-Coefficient Partially Linear Models," *Bernolli*, Vol. 11, No. 6, 2005, pp. 1031-1057.

[7] J. Fan, C. Zhang and J. Zhang, "Generalized Likelihood Ratio Statistics and Wilks Phenomenon," *The Annals of Statistics*, Vol. 29, No. 1, 2001, pp. 153-193.

[8] J. Fan and W. Zhang, "Statistical Estimation in Varying Coefficient Models," *The Annals of Statistics*, Vol. 27, No. 5, 1999, pp. 1491-1518.

[9] J. Fan and J. Zhang, "Two-Step Estimation of Functional Linear Models with Application to Longitudinal Data," *Journal of Royal Statistical Association B*, Vol. 62, No. 2, 2000, pp. 303-322.

[10] H. Wang and Y. Xia, "Shrinkage Estimation of The Varying Coefficient Model," *Journal of the American Statistical Association*, Vol. 104, No. 486, 2009, pp. 747-757.

[11] T. Hastie and R. Tibshirani, "Varying-Coefficient Models," *Journal of Royal Statistical Association B*, Vol. 55, No. 4, 1993, pp. 757-796.

[12] H. Koul, V. Susarla and J. Van Ryzin, "Regression Analysis with Randomly Right Censored Data," *The Annals of Statistics*, Vol. 9, No. 6, 1981, pp. 1276-1288.

[13] S. Leurgans, "Linear Models, Random Censoring and Synthetic Data," *Biometrika*, Vol. 74, No. 2, 1987, pp. 301-309.

[14] X. Luo, Z. Yang and Y. Zhou, "Varying-Coefficient Regression Models with Censored Data," *Acta Mathematicae Applicatae Sinica*, Vol. 29, No. 3, 2006, pp. 415-427.

[15] Z. Zheng, "A Class of Estimators of the Parameters in Linear Regression with Censored Data," *Acta Mathematicae Applicatae Sinica*, Vol. 3, No. 3, 1987, pp. 231-241.

[16] Z. Zheng, "Strong Consistency of Nonparametric Regression Estimates with Censored Data," *Journal of Mathematical Research and Exposition*, Vol. 8, No. 4, 1988, pp. 307-313.

[17] Y. Zhou and H. Liang, "Statistical Inference for Semiparametric Varying-Coefficient Partially Linear Models with Error-Prone Linear Covariates," *The Annals of Statistics*, Vol. 37, No. 1, 2009, pp. 427-458.

### Appendix

To derive the asymptotic distribution of  $\lambda_n$  under  $H_0$ , we need the following conditions.

**C1.** The marginal density  $f(u)$  of  $U$  is Lipschitz continuous and bounded away from 0.  $U$  has a bounded support  $\Omega$ .

**C2.**  $\alpha(u)$  has the continuous second derivative.

**C3.** The function  $K(t)$  is symmetric and bounded. Further, the function  $t^3K(t)$  and  $t^3K'(t)$  are bounded and  $\int t^4K(t)dt < \infty$ .

**C4.**  $E(\mathbf{X}\mathbf{X}^T\mathbf{X}\mathbf{X}^T\sigma^{*4}(X,U)|U=u) < \infty$ .

**C5.**  $\mathbf{X}$  is bounded and the  $p \times p$  matrix  $E(\mathbf{X}\mathbf{X}^T|U=u)$  is invertible for each  $u \in \Omega$ .  $E^{-1}(\mathbf{X}\mathbf{X}^T|U=u)$  and  $E(\mathbf{X}\mathbf{X}^T\sigma^{*2}(X,U)|U=u)$  are both Lipschitz continuous.

**Remark** Conditions (C1)-(C5) are standard conditions, which are commonly used in varying-coefficient regression model (see Fan, J. and Huang T, 2005 and Luo et al 2006).

**Lemma 1.** Suppose that  $f_U(\cdot)$  is positive and continuous on a compact interval  $[a,b]$ , and  $k_n \rightarrow \infty$

$$D_2 = \sum_{i=1}^n (\hat{\alpha}(U_i, \hat{\phi}_1, \hat{\phi}_2) - \alpha(U_i))^T \mathbf{X}_i \mathbf{X}_i^T (\hat{\alpha}(U_i, \hat{\phi}_1, \hat{\phi}_2) - \alpha(U_i)) - 2 \sum_{i=1}^n \varepsilon_i^* (\hat{\alpha}(U_i, \hat{\phi}_1, \hat{\phi}_2) - \alpha(U_i))^T \mathbf{X}_i,$$

$$= \left\{ \sum_{i=1}^n (\mathbf{e}_n(U_i) - \mathbf{R}_n(U_i))^T \mathbf{X}_i \mathbf{X}_i^T (\mathbf{e}_n(U_i) - \mathbf{R}_n(U_i)) - 2 \sum_{i=1}^n \varepsilon_i^* (\mathbf{e}_n(U_i) - \mathbf{R}_n(U_i))^T \mathbf{X}_i \right\} (1 + o_p(1)).$$

The proof will be completed by showing the following four step.

- (1)  $D_1 = O_p(1)$ ,
- (2)  $-\sqrt{h}D_2 = \frac{D}{\sqrt{h}} + W(n) + o_p(1)$ ,
- (3)  $W(n) = \frac{\sqrt{h}}{n} \sum_{j \neq i} \varepsilon_j^* \varepsilon_i^* [2K_h(U_j - U_i) - K_h * K_h(U_j - U_i)] \mathbf{X}_j^T \Gamma^{-1}(U_i) \mathbf{X}_i \xrightarrow{L} N(0, V)$ ,
- (4)  $\frac{RSS_1}{n} = E\sigma^{*2}(U, X) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{nh}\right)$ ,

with

$$D = [2K(0) - K * K(0)] \int_{\Omega} tr(\Gamma^*(u)\Gamma(u)^{-1}) du - \frac{1}{nh} K^2(0) E\left[(X^T \Gamma(U)^{-1} X)^2 \sigma^{*2}(X, U)\right],$$

$$V = 2 \int [2K(x) - K * K(x)]^2 dx \int_{\Omega} tr(\Gamma^*(u)\Gamma(u)^{-1})^2 du.$$

It follows from Lemma 7.1 in Fan et al(2001) that

$$E\left[\left(\varepsilon^{*T} P_{X_D} \varepsilon^*\right)^2 | (X_1, U_1), \dots, (X_n, U_n)\right] \leq ctr(P_{X_D}^2) + c \left[tr(P_{X_D})\right]^2 = p(p+1)c$$

which implies (1). The proofs of (2) and (3) are the same as the proof of Theorem 5 in Fan et al (2001). The

such that  $k_n/n \rightarrow 0$ . Then,  $\hat{h}_{k_n} = \frac{k_n}{nf_U(u)}(1 + o_p(1))$  uniformly in  $u \in [a, b]$ .

**Proof.** This can be shown by the proof of Theorem 5.1. in Fan and Gijbels (1994).

**Lemma 2.** Let  $\hat{\alpha}$  be the local linear estimator defined in section 2. Then, under condition (C1)-(C5), uniformly for  $u_0 \in \Omega$ ,

$$\hat{\alpha}(u_0, \hat{\phi}_1, \hat{\phi}_2) - \alpha(u_0) = (e_n(u_0) - R_n(u_0))(1 + o_p(1)),$$

where  $e_n(u_0) = \frac{\Gamma(u_0)^{-1}}{nh} \sum_{i=1}^n \varepsilon_i^* \mathbf{X}_i K((U_i - u_0)/h)$ ,

$$R_n(u_0) = 1/2h^2 v_2 \alpha_0''(u_0) \text{ with } v_2 = \int u^2 K^2(u) du.$$

**Proof.** This follows immediately from the result obtained by Luo et al (2006).

**Proof of Theorem 1.** Let  $c$  denote a generic constant. Then, under  $H_0$ ,

$$RSS_0 - RSS_1 = -D_1 - D_2,$$

where  $D_1 = \varepsilon^{*T} P_{X_D} \varepsilon^*$ ,  $X_D$  is the design matrix with the  $i$  th row  $\mathbf{X}_i^T$  ( $i=1, \dots, n$ ) and  $P_{X_D}$  is the projection matrix of  $X_D$  and

details are omitted. The last step follows from  $RSS_1 = \sum_{i=1}^n \varepsilon_i^{*2} + D_2$ .