Statistica Sinica **13**(2003), 763-781

CHECKING THE ADEQUACY OF A PARTIAL LINEAR MODEL

Li-Xing Zhu and Kai W. Ng

Chinese Academy of Sciences and The University of Hong Kong

Abstract: A partial linear model is a model where the response variable depends on some covariates linearly and on others nonparametrically. In this article, we construct an empirical process-based test for examining the adequacy of partial linearity of model. A re-sampling approach, called random symmetrization (RS), is applied to obtain the approximation to the null distribution of the test. The procedure is easy to implement. A simulation study is carried out and application to an example is made.

Key words and phrases: Conditional distribution, empirical process, partially linear model, re-sampling.

1. Introduction

The partial linear model has received considerable attention. It is written as:

$$Y = \beta' X + g(T) + \varepsilon,$$

where X is a d-dimensional random vector, T is a 1-dimensional random variable, β is an unknown parameter vector of d-dimension, $g(\cdot)$ is an unknown measurable function, and the conditional expectation of ε given (T, X) is zero. Without loss of generality, we assume X has zero mean. There are many proposals in the literature for the estimation of β and g. Among them are Cuzick (1992), Engle, Granger, Rice and Weiss (1986), Mammen and van der Geer (1997) and Speckman (1988).

In this paper, we consider testing

$$H_0 : E(Y|X = \cdot, T = \cdot) = \beta' \cdot + g(\cdot) \ , \quad \text{for some } \beta \text{ and } g, \tag{1.1}$$

against $H_1 : E(Y|X = \cdot, T = \cdot) \neq \beta' \cdot + g(\cdot)$ for any β and g.

For testing H_0, Whang and Andrews (1993) and Yatchew (1992) used sample splitting to recommend *ad hoc* methods. Fan and Lin (1996) employed a kernel smoother to estimate the conditional expectation of residuals given (X, T), and constructed a test with a limiting normal null distribution. They reported asymptotic results but no simulation to demonstrate performance. Their test, however, may have problems due to the inefficiency of kernel estimation for high dimensional data.

In the literature, there are several approaches available for constructing test statistics: for parametric models, Dette (1999) suggested a test based on the difference of variance estimators; Eubank and Hart(1992) studied a test of score type; Eubank and LaRiccia (1993) proposed a method of variable selection; Härdle and Mammen (1993) considered a test statistic of the difference between parametric and nonparametric fits; Stute, Manteiga and Quibdinil (1998) applied a test based on a residual-marked process; Stute, Theis and Zhu (1998) proposed an innovation approach so as to determine p-values conveniently; Fan and Huang (2001) suggested an adaptive Neyman test. Hart (1997) contains fairly comprehensive references.

Our test is based on a residual-marked process. The main reasons that we use this approach are as follows: Since our setting has a multivariate covariate, the Härdle and Mammen (1993) test may suffer dimensionality problems because local smoothing for the nonparametric fit is involved. As for the adaptive Nyeman test (Fan and Huang (2001)), we have not yet seen whether it is asymptotically distribution free when the distribution of error is not normal. Moreover, its power performance depends on the smoothness of $\varepsilon_j = y_j - \beta' x_j - g(t_j)$ as a function of j. As Fan and Huang (2001) point out, achieving sufficient smoothness with multivariate predictors is very challenging.

On the other hand, although the test based on a residual-marked process has some drawbacks, mainly the insensitiveness to the alternatives of oscillatory regression functions, it still shares some desirable features: it is consistent for all global alternatives; it is able to detect local alternatives of order arbitrarily close to $n^{-1/2}$; it is asymptotically distribution-free; only one-dimensional nonparametric function estimation is required for the computation.

As is known, $n^{-1/2}$ is the best achievable rate for lack of fit tests. The optimal rate of the adaptive Neyman test is $O(n^{-s/(2s+1)}(\log\log n)^{s/(4s+1)})$ for some $s > 0$ (see Spokoiny (1996) or Fan and Huang (2001)). The fourth feature is particularly desirable for multivariate regression.

When the exact or limiting null distribution of a test statistic is intractable for computing p-values, one frequently resorts to the use of bootstrap approximations. In a parametric context, bootstrap approximations can maintain significance level and have good power performance; see Härdle and Mammen (1993) and Stute, Manteiga and Quibdinil (1998). Our circumstances are, however, more complicated. The classical bootstrap approximation has been shown to be inconsistent in the parametric case (see, e.g., Stute,Manteiga and Quibdinil

(1998)). Even for the wild bootstrap approximation, consistency is not clear because, in a related work of heteroscedasticity checking, it was shown not to work. See Zhu, Fujikoshi and Naito (2001).

Here we apply a variant of the wild bootstrap approximation suggested by Zhu, Fujikoshi and Naito (2001). It is motivated by the "Random Symmetrization" of Pollard (1984), and also by similar techniques of Dudley (1978) and Giné and Zinn (1984). Hence we call it the random symmetrization (RS) method. A related work is Zhu, Ng and Jing (2001).

The article is organized in the following way. In the next section, we construct the test and study its asymptotic behavior. In Section 3, the consistency of the RS approximation is presented. Some simulation results are reported, and application to an example is made, in Section 4. All proofs are postponed to Section 5.

2. A Test Statistic and Its Limiting Behavior

2.1. Motivation and construction

We first describe the construction of a residual-marked cusum process. For any weight function $w(T)$, let

$$U(T,X) = (X - E(X|T)),\ \ V(T,Y) = (Y - E(Y|T)),\ \ S = E(UU'w^2(T)), \quad (2.1)$$

$$\beta = S^{-1}E[U(T,X)V(T,Y)w^2(T)],\quad \gamma(t) = E(Y|T=t), \quad (2.2)$$

for a positive definite matrix S. Fan and Li (1996) considered the density function of T as a weight function $w(\cdot)$. A constant weight whose support is (a,b) with $0 < a < b < 1$ is also often applied. Note that H_0 is true if and only if

$$E[Y - \beta'U(T,X) - \gamma(T)]w(T)I(T \le t, X \le x) = 0, \qquad \text{for all } t, x, \quad (2.3)$$

where $X \le x$ means that each component of X is less than or equal to the corresponding component of x, similarly for $T \le t$. Let $\{(t_1,x_1,y_1),\ldots,(t_n,x_n,y_n)\}$ be a set of observations. The empirical version of the LHS of (2.3) is, letting $\hat\varepsilon_j = y_j - \hat\beta'\hat U(t_j,x_j) - \hat\gamma(t_j)$,

$$R_n(t,x) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \hat\varepsilon_j w(t_j) I(t_j \le t, x_j \le x), \quad (2.4)$$

where $\hat\beta$, $\hat U = X - \hat E(X|T)$ and $\hat\gamma$ are the estimators of β, U and γ. The proposed test statistic is defined as

$$CV_n = \int (R_n(T,X))^2 d\,F_n(T,X), \quad (2.5)$$

where F_n is the empirical distribution based on $\{(t_1, x_1), \ldots, (t_n, x_n)\}$. We reject the null hypothesis for large values of CV_n.

Note that CV_n is not a scale-invariant statistic. Usually a normalizing constant is needed, say the estimator of the limiting variance, when the limiting null distribution of the test is used for p-values, Fan and Li (1996) and Fan and Huang (2001). However, it is not easy to choose a proper variance estimator which does not weaken the power performance of test. We do not need such a normalizing constant because in the RS approximation, it is constant for given $(t_i, x_i, y_i)'s$ and does not have any impact on the conditional distribution of the RS test statistic. Details are in Section 3.

2.2. Estimation of β and γ

For $i = 1, \ldots, n$, let $\hat{f}_i(t_i) = (1/n)\sum_{j\neq i}^n k_h(t_i - t_j)$ with

$$\hat{E}_i(X|T = t_i) = \frac{1}{n}\sum_{j\neq i}^n x_j k_h(t_i - t_j)/\hat{f}_i(t_i),$$

$$\hat{E}_i(Y|T = t_i) = \frac{1}{n}\sum_{j\neq i}^n y_j k_h(t_i - t_j)/\hat{f}_i(t_i),$$

$$\hat{U}(t_i, x_i) = x_i - \hat{E}_i(X|T = t_i),\ \hat{V}(t_i, y_i) = y_i - \hat{E}_i(Y|T = t_i),$$

$$\hat{S} = \hat{E}(\hat{U}\hat{U}'w^2(T)) = \frac{1}{n}\sum_{j=1}^n \hat{U}(t_j, x_j)\hat{U}(t_j, x_j)'w^2(t_j),$$

where $k_h(t) = (1/h)K(t/h)$ and $K(\cdot)$ is a kernel function satisfying the condition in the appendix. The resulting estimators are

$$\hat{\beta} = (\hat{S})^{-1}(1/n)\sum_{j=1}^n \hat{U}(t_j, x_j)\hat{V}(t_j, y_j)w^2(t_j), \quad \hat{\gamma}(t_i) = \hat{E}_i(Y|T = t_i). \tag{2.6}$$

Under the conditions in the appendix, we can follow along the lines of Schick (1996) to derive the following proposition:

Proposition. *Under conditions* 1−6 *in the appendix,*

$$\sqrt{n}(\hat{\beta} - \beta) = S^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^n U(t_j, x_j)\varepsilon_j w^2(t_j) + O_p([\frac{1}{h\sqrt{n}} + h^2\sqrt{n}]^{1/2}) \tag{2.7}$$

converges in distribution to $N(0, S^{-1}E[(U(T, X)U(T, X)'w^4(T)\varepsilon^2]S^{-1})$, *and for any subset* $[a, b]$ *with* $0 < a < b < 1$,

$$\sup_{a\le t\le b} |\hat{\gamma}(t) - \gamma(t)| = O_p(\frac{1}{\sqrt{nh}} + h). \tag{2.8}$$

For details of the proof, refer to Technical Report 342 of the Department of Statistics and Actuarial Science, the University of Hong Kong.

2.3. Asymptotic properties of the test

We now state the asymptotic properties of R_n and CV_n. Let

$$\begin{aligned}&J(T,X,Y,\beta,U,S,F(X|T),t,x,)\\&=\varepsilon w(T)\Big\{I(T\le t,X\le x)-E\Big[I(T\le t,X\le x)U(T,X)'w(T)\Big]S^{-1}U(T,X)\\&\qquad -F(X|T)I(T\le t)\Big\},\end{aligned}$$

where $F(x|T)$ is the conditional distribution of X given T.

Theorem 2.1. *Under conditions* 1−6 *in the appendix, under* H_0*,*

$$R_n(t,x)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}J(t_j,x_j,y_j,\beta,U,S,F(x_j|t_j),t,x,)+o_p(1)$$

converges in distribution to R *in the Skorokhod space* $D[-\infty,+\infty]^{(d+1)}$*, where* R *is a centered continuous Gaussian process with the covariance function*

$$\begin{aligned}&E(R(t_1,x_1)(R(t_2,x_2))\\&=E(J(T,X,Y,\beta,U,S,F(X|T),t_1,x_2,)J(T,X,Y,\beta,U,S,F(X|T),t_2,x_2,)).\end{aligned}\tag{2.9}$$

Therefore, CV_n *converges in distribution to* $CV:=\int R^2(T,X)dF(T,X)$ *with* $F(\cdot,\cdot)$ *being the distribution function of* (T,X)*.*

We now investigate how sensitive the test is to alternatives. Consider a sequence of models indexed by n

$$E(Y|X,T)=\alpha+\beta'X+g(T)+g_1(T,X)/\sqrt{n}.\tag{2.10}$$

Theorem 2.2. *In addition to the conditions of Theorem* 2.1*, assume that* $g_1(T,X)$ *has zero mean and satisfies the condition: there exists a neighborhood of the origin,* U*, and a constant* $c>0$ *such that, for any* $u\in U$*,*

$$|E(g_1(T,X)|T=t+u)-E(g_1(T,X)|T=t)|\le c|u|\quad\textit{for all } t \textit{ and } x.$$

Then, under the alternative (2.10)*,* R_n *converges in distribution to* $R+g_{1*}$ *where*

$$\begin{aligned}g_{1*}(t,x)=&E\Big\{[g_1(T,X)-E(g_1(T,X)|T)]w(T)I(T\le t,X\le x)\Big\}\\&-E\{U(T,X)'(g_1(T,X)-E(g_1(T,X)|T))w^2(T)\}S^{-1}\\&\times E\Big\{U(T,X)w(T)I(T\le t,X\le x)\Big\}\end{aligned}$$

is a non-random shift function. Thus CV_n *converges in distribution to* $\int(B(T,X) + g_{1*}(T,X))^2 dF(T,X)$.

From the expression for g_{1*}, we realize that it cannot vanish unless $g_1(T,X) = \beta' X$. Hence the test CV_n is capable of detecting local alternatives arbitrarily close to $n^{-1/2}$ from the null. From the proof of the theorem in the appendix, it is easy to see that the test is consistent against any global alternative such that $g_1(T,X)w(T)$ is not a constant function with respect to $T \in [a,b]$ and X.

Our test can also detect alternatives with a departure function $c_n g(x/b_n)$, where c_n and b_n converge to zero and $n^{1/2}(c_n \times b_n) \to \infty$. This can be shown by the argument of Theorem 2.2 when some more regularity conditions on function $G(\cdot)$ are assumed. On the other hand, our test cannot detect any alternative. Further study of the optimality of the test against certain classes of alternatives is merited.

3. A Re-sampling Approximation

Let

$$J_1(T,X,Y,t,x,\beta) = \varepsilon w(T) I(T \le t, X \le x),$$
$$J_2(T,X,Y,t,x,U,S) = \varepsilon w(T) E[U(T,X)' w(T) I(T \le t, X \le x)] S^{-1} U(T,X),$$
$$J_3(T,X,Y,t,x,\beta,F_{X|T}) = \varepsilon w(T) F(X|T) I(T \le t).$$

Then $J(T,X,Y,t,x,\beta,U,S,F_{X|T}) = J_1(T,X,Y,t,x) - J_2(T,X,Y,U,S,t,x) - J_3(T,X,Y,\beta,F(X|T),t,x)$. From Theorem 2.1, we have that, asymptotically, $R_n(t,x) = (1/\sqrt{n}) \sum_{j=1}^n J(t_j,x_j,y_j,\beta,U,S,F_{x_j|t_j},t,x)$.

The procedure is as follows (see e.g., Zhu, Fujikoshi and Naito (2001)).

- *Step* 1. Generate random variables $e_i, i = 1,\ldots,n$, independent with mean zero and variance one. Let $E_n := (e_1,\ldots,e_n)$ and define the conditional counterpart of R_n as

$$R_n(E_n,t,x) = \frac{1}{\sqrt{n}} \sum_{j=1}^n e_j J(t_j,x_j,y_j,\hat\beta,\hat U,\hat S,\hat F_{x_j|t_j},t,x,), \tag{3.1}$$

  where $\hat\beta$, $\hat U$, $\hat S$, $\hat F$ are consistent estimators of the unknowns in R_n. The resultant conditional test statistic is

$$CV_n(E_n) = \int (R_n(E_n))^2 F_n(t,x). \tag{3.2}$$

- *Step* 2. Generate m sets of E_n, say $E_n^{(i)}$, $i = 1,\ldots,m$, and get m values of $CV_n(E_n)$, $CV_n(E_n^{(i)})$, $i = 1,\ldots,m$.

- *Step* 3. The p-value is estimated by $\hat{p} = k/(m+1)$ where k is the number of $CV_n(E_n^{(i)})$'s which are larger than or equal to CV_n. Reject H_0 when $\hat{p} \le \alpha$ for a designated level α.

The following result states the consistency of the approximation.

Theorem 3.1. *Under either H_0 or H_1 and the conditions in Theorem* 2.1*, we have that, for almost all sequences $\{(t_1, x_1, y_1), \ldots, (t_n, x_n, y_n), \ldots\}$, the conditional distribution of $R_n(E_n)$ converges to the limiting null distribution of R_n.*

Remark 3.1. The conditional distribution of $CV_n(E_n)$ serves for determining p-values of the test, we naturally hope that the conditional distribution can well approximate the null distribution of the test statistic no matter whether data are from the hypothetized or the alternative model. On the other hand, as we do not know the underlying model of the data when the re-sampling method is applied, we take the risk that under the alternative the conditional distribution is far away from the null distribution of the test. If so, the determination of the p-values can be inaccurate and damage the power performance. Theorem 3.1 indicates that the conditional distribution based on the RS approximation may avoid this trouble.

Remark 3.2. Here is why we need not choose a normalizing constant in constructing the test statistic. In view of Theorem 2.1, we know that a normalizing constant could be $C_n = \sup_{t,x} 1/n \sum_{j=1}^{n} \Big(J(t_j, x_j, y_j, \hat{\beta}, \hat{U}, \hat{S}, \hat{F}_{x_j|t_j}, t, x,)\Big)^2$, the supremum of the sample variance of $J(T, X, Y, \beta, U, S, F_{X|T}, t, x,)$ over t and x. Looking at (3.1), we realize that this is constant when the $(t_i, x_i, y_i)'$s are given. For determining p-values, CV_n/C_n associated with $CV_n(E_n)/C_n$ is equivalent to CV_n associated with $CV_n(E_n)$.

4. Simulation Study and Application

In the simulations we conducted, the underlying model was

$$y = \beta x + bx^2 + (t^2 - 1/3) + \sqrt{12}(t - 1/2)\varepsilon, \tag{4.1}$$

where t is uniformly distributed on [0,1], x and ε are random variables. We considered the following four cases, as suggested by a referee. (1). Uni-Uni: both x and ε are uniformly on [-0.05, 0.5]; (2). Nor-Uni: standard normal x and uniform ε on [-0.5, 0.5]; (3). Nor-Nor: standard normal x and standard normal ε; (4). Uni-Nor: uniform x on [-0.5, 0.5] and standard normal ε. The empirical powers of these four cases are plotted in Figure 1. In the simulations, we chose $\beta = 1$ and $b = 0.0, 0.5, 1.0, 1.5$ and 2.0 for showing the power performance at different alternatives. Note that $b = 0.0$ corresponds to H_0. The sample size was

100 and the nominal level was 0.05. The experiment was performed 3000 times. We chose $K(t) = (15/16)(1-t^2)^2 I(t^2 \leq 1)$ as the kernel function, it has been used by, for example, Härdle (1988) for estimation and Härdle and Mammen (1993) for hypothesis testing. Another referee pointed out that the choice of bandwidth is one of major concerns in hypothesis testing. The difficulty with the theorem's treatment is that it does not allow a data-driven choice for h. Indeed, it is not even clear how h should be selected in this setting. Fan and Li (1996) did not discuss this issue at all. Gozalo and Linton (2001), in a related work, employed generalized cross-validation (GCV) to select the bandwidth without arguing its use. Eubank and Hart (1993) state that with homoscedastic errors GCV is useful, while with heteroscedastic errors its usefulness is not clear. Selecting a bandwidth in hypothesis testing is still an open problem and is beyond the scope of this paper. In our simulation, in order to obtain some insight on how the bandwidth should be chosen, we combined GCV with a grid search. We first computed the average value of h, h_{egcv}, selected by GCV over 1000 replications, then we performed a grid search over $[h_{egcv}$-1, h_{egcv}+1]. For the cases with uniform error, $h = 0.30$ worked best, with $h = 0.57$ best for the cases with normal error. The size is close to the target value of 0.05.

We considered a comparison with Fan and Li's (1996) test (FL). Since the FL test involves kernel estimation with all covariates including t, we used a product kernel, each factor of which was $K(t) = (15/16)(1 - t^2)^2 I(t^2 \leq 1)$. In our initial simulation for the FL test, we were surprised to find that the test had almost has no power. We found that the estimate of variance being used has severe influence on power performance since, under the alternatives, its value gets fairly large. Based on this observation, we used the estimate of variance under H_0, with some constant adjustment so as to maintain the significance level. The results of the power are reported in Figure 1. Since we know of no other tests for partial linearity except the FL test, we also included a comparison with the adaptive Neyman test of Fan and Huang (2001), who reported that the test was able to detect nonparametric deviations from a parametric model with Gaussian error. The estimate of the variance is also the one under the null hypothesis in Figure 1, Adj-FL and Adj-FH stand, respectively, for the FL test and Fan and Huang's test with the adjustment of variance estimation.

Looking at Figure 1(1)$-$(4), for uniform x, CV_n has higher power than Adj-FL and Adj-FH. The adaptive Neyman test Adj-FH works well with the normal covariate x, see Figure 1(2) and (3), while our test does not perform well in this case. After the adjustment, Adj-FL is very sensitive to the alternative in the Nor-Uni case. It seems that there is no uniformly best test here.

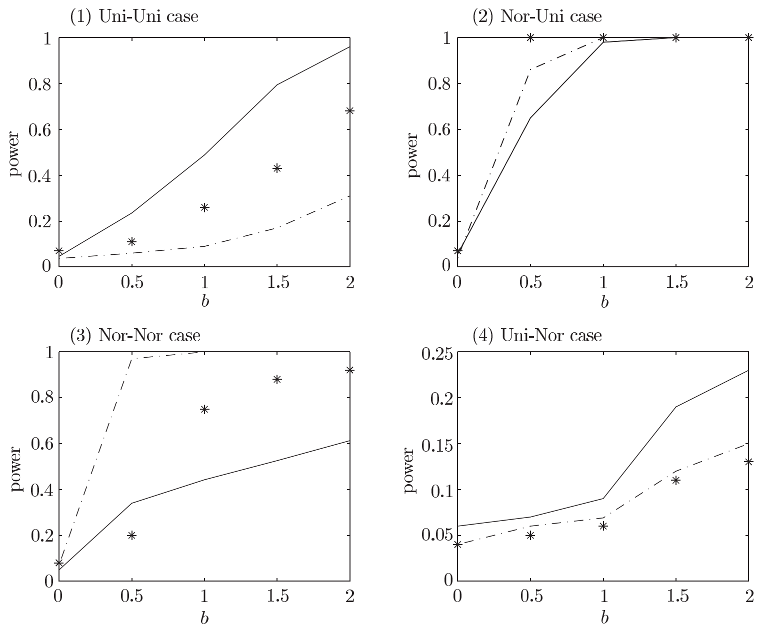

Figure 1. In all plots, the solid line is for the test CV_n with a combined search of GCV and grid points; the dashdot is for the ADJ-FH test; the star is for the ADJ-FL test.

Example. The data are the depths and locations of $n = 43$ earthquakes occurring near the Tonga trench between January 1965 and January 1966 (see Sykes, Isacks and Oliver (1969)). The variable X_1 is the perpendicular distance in hundreds of kilometers from a line that is approximately parallel to the Tonga trench. The variable X_2 is the distance in hundreds of kilometers from an arbitrary line perpendicular to the Tonga trench. The response variable Y is the depth of the earthquake in hundreds of kilometers. Under the plate model, the depths of the earthquakes will increase with distance from the trench and the scatter plot of the data in Figure 2 shows this to be the case. Our purpose is to check whether the plate model is linear or not. Looking at Figure 2 we find that the plots of Y against X_1 indicate an apparent linear relation with heteroscedasticity, while that between Y and X_2 is not very clear. We find that a linear model is tenable for Y against X_1 with the fitted value $\hat{Y} = -0.295 + 0.949X_1$, while the linear model for Y against X_2 is rejected. If we try a linear model for Y against both X_1 and X_2, $Y = \hat{\beta}_0 + \hat{\beta}^\tau X$ where $X = (X_1, X_2)^\tau$, Figure 2 (5) and (6) would support the linearity between Y and X, and the residual plot against $\hat{\beta}^\tau X$ shows a similarity to that of Y against X_1 in Figure 2 (2). This finding may be explained as a greater impact on Y by X_1 than by X_2. However in Figure 2 (4), we can see that there is some curved structure between the residuals and X_2. The effect of X_2 is not negligible. These observations lead to a more complicated modelling.

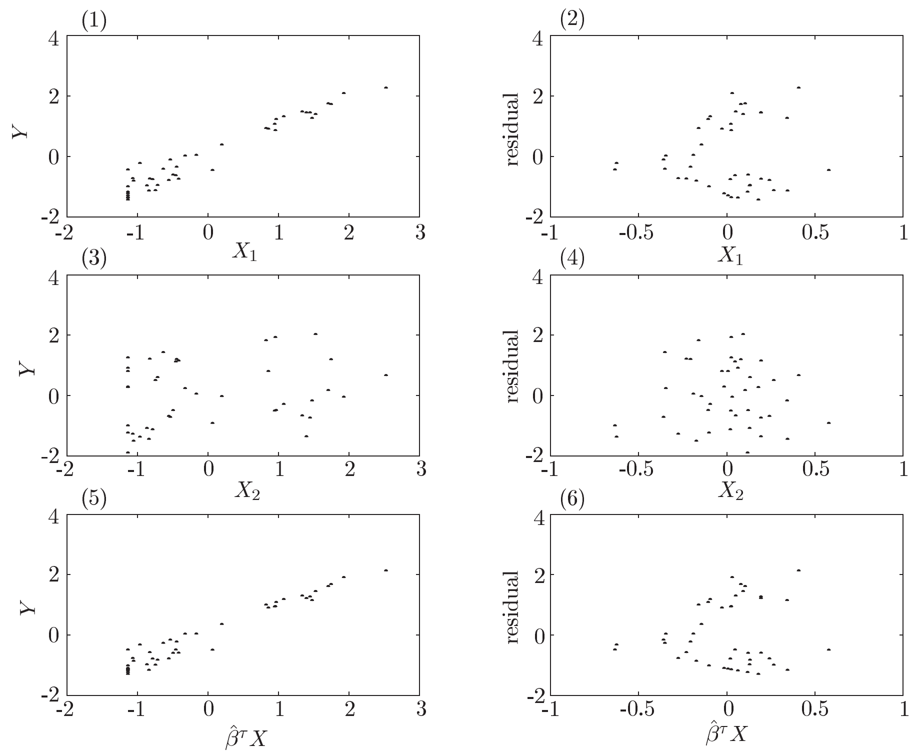

Figure 2. (1), (3) and (5) are scatter plots of Y against X_1, X_2 and $\hat{\beta}^\tau X$, where $\hat{\beta}$ is the least squares estimator of β; (2), (4) and (6) are the residual plots against X_1, X_2 and $\hat{\beta}^\tau X$ when a linear model is used.

A partial linear model $Y = \beta_0 + \beta_1 X_1 + g(X_2) + \varepsilon = \beta_0 + \beta_1 U + r(X_2) + \varepsilon$, with $U = X_1 - E(X_1|X_2)$, provides some reasonable interpretation. Looking at Figure 3 (b), we would have $E(X_1|X_2)$ a nonlinear function of X_2. Checking Figure 3 (c), $(Y - \beta_0 - \beta U)$ should be nonlinear. The residual plot against X_2 in this modelling shows that there is no clear indication of relation between the residual and X_2. Using the test suggested in the present paper, we have $T_n = 0.009$ and the p-value is 0.90. A partial linear model is tenable.

A. Appendix

A.1. Assumptions

(1) Write the first derivative of $E(Y|T = t)$ as $E^{(1)}(Y|T = t)$. Assume that $E^{(1)}(Y|T = t)$, $E(X|T = t)$ and the conditional distribution function of X given

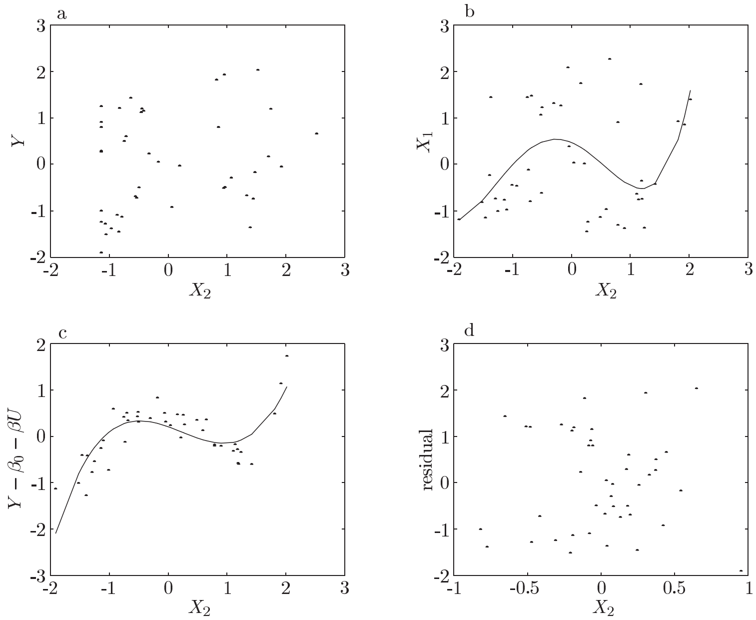

FIgure 3. (a). Scatter plot of Y against X_2; (b). Scatter plot of X_1 against X_2 and the fitted curve $\hat{E}(X_1|X_2)$ (the solid line); (c). Scatter plot of $Y - \beta_0 - \beta U$ where $U = X_1 - E(X_1|X_2)$ and the fitted curve of $E(Y|X_2)$ (the solid line); (d). The residual plot against X_2 when the data are fitted by partially linear model $Y = \beta_0 + \beta X_1 + g(X_2)$.

$T = t$, $F(x|t)$ say, all satisfy the following condition: there exists a neighborhood of the origin, say U, and a constant $c > 0$ such that for any $u \in U$ and all t and x,

$$\begin{aligned} |E(X|T = t + u) - E(X|T = t)| &\le c|u|; \\ |E^{(1)}(Y|T = t + u) - E^{(1)}(Y|T = t)| &\le c|u|; \\ |F(x|t + u) - F(x|t)| &\le c|u|. \end{aligned} \tag{A.1}$$

(2) $E|Y|^4 < \infty$ and $E|X|^4 < \infty$.
(3) The continuous kernel function $K(\cdot)$ satisfies:
  (a) the support of $K(\cdot)$ is the interval $[-1, 1]$;
  (b) $K(\cdot)$ is symmetric about 0;
  (c) $\int_{-1}^{1} K(u)du = 1$, and $\int_{-1}^{1} |u|K(u)du \neq 0$.
(4) As $n \to \infty$, $\sqrt{n}h^2 \to 0$ and $\sqrt{n}h \to \infty$.
(5) $E(\varepsilon^2|T = t, X = x) \le c_1$ for some c_1 and all t and x.
(6) The weight function $w(\cdot)$ is bounded and continuous on its support set $[a, b]$, $-\infty < a < b < \infty$, on which $f(\cdot)$ is bounded away from zero.

Remark A.1. Conditions (1) and (3) are typical for showing convergence rates of nonparametric estimates. Condition (2) is necessary for the asymptotic normality of a least squres estimate. Condition (4) ensures the convergence of the test statistic. Condition (6) is to avoid the boundary effect when a nonparametric smoothing is applied.

A.2. Proofs, Section 2

Proof of Theorem 2.1. The proposition in Subsection 2.2 and some elementary calculation show that

$$\begin{aligned} R_n(t,x) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j w(t_j) I(t_j\le t, x_j\le x) \\ &\quad -E(U(T,X)'w(T)I(T\le t, X\le x)S^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}U(t_j,x_j)\varepsilon_j w(t_j) \\ &\quad -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\hat{g}(t_j)w(t_j)I(t_j\le t, x_j\le x) \\ &\quad +\frac{1}{\sqrt{n}}\sum_{j=1}^{n}g(t_j)w(t_j)I(t_j\le t, x_j\le x)+O_p(\frac{1}{h\sqrt{n}}+\sqrt{n}h^2) \\ &=: I_1(t,x)-I_2(t,x)-I_3(t,x)+I_4(t,x)+O_p(\frac{1}{h\sqrt{n}}+\sqrt{n}h^2). \quad \text{(A.2)} \end{aligned}$$

The convergence of I_1 and I_2 follows standard empirical process theory, see, e.g., Pollard (1984, Chap. VII). We now prove that $I_3 - I_4$ converges in distribution to a Gaussian process. Deal with I_3. By some calculation we can derive that, letting $\hat{r}_i(t_i) = (1/n)\sum_{j\neq i}^{n}(y_j - \beta' x_j)k_h(t_i - t_j)$ and invoking the **Proposition**, for $a \le t_j \le b$,

$$\begin{aligned} \hat{g}(t_j) &= \frac{\hat{r}_j(t_j)}{\hat{f}_j(t_j)} = \frac{\hat{r}_j(t_j)}{f(t_j)} + \frac{r(t_j)}{f(t_j)}\frac{f(t_j)-\hat{f}_j(t_j)}{f(t_j)} \\ &\quad +\frac{r(t_j)}{f(t_j)}\frac{(f(t_j)-\hat{f}_j(t_j))^2}{\hat{f}_j(t_j)f(t_j)} + \frac{\hat{r}_j(t_j)-r(t_j)}{f(t_j)}\frac{f(t_j)-\hat{f}_j(t_j)}{\hat{f}_j(t_j)} \\ &= \frac{\hat{r}_j(t_j)}{f(t_j)} + g(t_j)\frac{f(t_j)-\hat{f}_j(t_j)}{f(t_j)} + O_p(\frac{1}{hn}+h^2), \end{aligned}$$

where $g(t_j) = r(t_j)/f(t_j)$. Hence

$$I_3(t,x) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\hat{r}_j(t_j)}{f(t_j)}w(t_j)I(t_j\le t, x_j\le x)$$

$$
\begin{aligned}
&-\frac{1}{\sqrt{n}}\sum_{j=1}^{n} g(t_j)\frac{\hat{f}_j(t_j)}{f(t_j)}w(t_j)I(t_j \le t, x_j \le x)\\
&+\frac{1}{\sqrt{n}}\sum_{j=1}^{n} g(t_j)w(t_j)I(t_j \le t, x_j \le x) + O_p(\frac{1}{h\sqrt{n}} + \sqrt{n}h^2)\\
=:\ & I_{31}(t,x) - I_{32}(t,x) + I_{33}(t,x) + O_p(\frac{1}{h\sqrt{n}} + \sqrt{n}h^2).
\end{aligned}
$$

We now rewrite $I_{31}(t,x)$ as a U-statistic. Let $w_1(t) = w(t)/f(t)$ and $U_h(t_i, x_i, y_i; t_j, x_j, y_j; t, x) = [(y_i - \beta' x_i)w_1(t_j)I(t_j \le t, x_j \le x) + (y_j - \beta' x_j)w_1(t_i)I(t_i \le t, x_i \le x)]k_h(t_i - t_j)$. By the symmetry of $k_h(\cdot)$ we have, for fixed h (i.e. for fixed n), $hI_3(t,x) = (1/2n^{3/2})\sum_{j=1}^{n}\sum_{i\neq j}^{n} h\,U_h(t_i, x_i, y_i; t_j, x_j, y_j; t, x)$. For the sake of convenience, let $\eta_j = (t_j, x_j, y_j)$, and

$$
\begin{aligned}
&I'_{31}(\eta_1,\eta_2,t,x)\\
=&\frac{n}{n-1}h\,I_{31}(t,x) - E(h\,I_{31}(t,x)) - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{E[h\,U_h(\eta_1; t_j, x_j, y_j; t, x)]\\
&-E[h\,U_h(T,X,Y;T_1,X_1,Y_1;t,x)]\}.
\end{aligned}
$$

Note that $E(I'_{31}(\eta_1,\eta_2,t,x)) = 0$ and that the class $\mathcal{G}_n$ of functions consisting of $hU_h(\cdot,t,x) - E[hU_h(\eta_1;\eta_j t,x)]$ over all t and x is a Vapnik-Cervonenkis (VC) class of functions. Therefore $\mathcal{G}_n$ is P-degenerate with the envelope

$$
\begin{aligned}
G_n(\eta_1,\eta_2) = &\Big|[(y_1-\beta' x_1)w_1(t_2) + (y_2 - \beta' x_2)w_1(t_1)]|k((t_1-t_2)/h)\Big|\\
&+2\Big|E[(Y_1-\beta' Y_1)w_1(T_2) + (Y_2-\beta' X_2)w_1(T_1)\Big]|k(T_1-T_2)/h)\Big|\\
&+\Big|E[(Y_1-\beta' Y_1)w_1(t_2) + (y_2-\beta' x_2)w_1(T_1)\Big]|k(T_1-t_2)/h)\Big|.
\end{aligned}
$$

By Theorem 6 of Nolan and Pollard (1987, p.786), we have

$$
E\sup_x |\sum_{i,\ j} I'_3(\eta_i,\eta_j,t,x)| \le cE(\alpha_n + \gamma_n J_n(\theta_n/\gamma_n))/n^{-3/2},
$$

$$
\begin{aligned}
J_n(s) &= \int_0^s \log N_2(u, T_n, \mathcal{G}_n, G_n)d\ u,\\
\gamma_n &= (T_n G_n^2)^{1/2}, \qquad \alpha_n = \frac{1}{4}\sup_{g\in\mathcal{G}_n}(T_n g^2)^{1/2},\\
T_n g^2 &:= \sum_{i\neq j} g^2(\eta_{2i},\eta_{2j}) + g^2(\eta_{2i},\eta_{2j-1}) + g^2(\eta_{2i-1},\eta_{2j}) + g^2(\eta_{2i-1},\eta_{2j-1})
\end{aligned}
$$

for any function g, and $N_2(\cdot, T_n, \mathcal{G}_n, G_n)$ is the covering number of $\mathcal{G}_n$ under L_2 metric with the measure T_n and the envelope G_n. As $\mathcal{G}_n$ is a VC class, following the argument of Lemma II 2.25 of Pollard (1984, p.27) the covering number $N_2(uT_n/n^2G_n^2, T_n/n^2, \mathcal{G}_n, G_n)$ can be bounded by cu^{-w_1} for some positive c and w_1, both being independent of n and T_n. Furthermore in probability for large n, $T_nG_n^2 = O(hn^2\log^2 n)$ a.s. Hence $T_nG_n^2/n^2$ is smaller than 1 as $h = n^{-c}$ for some $c > 0$ and $N_2(u, T_n/n^2, \mathcal{G}_n, G_n) \le cu^{-w_1}$. Note that $N_2(u, T_n, \mathcal{G}_n, G_n) = N_2(u/n^2, T_n/n^2, \mathcal{G}_n, G_n)$. We can then show that $J_n(\theta_n/\gamma_n) \le J_n(1/4) = n^2 \int_0^{1/(4n^2)} \log N_2(u, T_n/n^2, \mathcal{G}_1, G)d\,u = -cn^2\int_0^{1/(4n^2)}\log u d\,u = c\log n$ and $\gamma_n^2 = T_nG_n^2 = O(hn^2\,\log^2 n)$ a.s. Therefore $E\sup_{t,x}|\sum_{i,j} I'_{31}(\eta_i,\eta_j,t,x)| \le c\sqrt{h/n}\log n$. Equivalently

$$I'_{31}(t,x) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{E[U_h(T,X,Y;t_j,x_j,y_j;t,x)] + O_p(\log n/\sqrt{nh}). \tag{A.3}$$

Consequently, noting $(n/(n-1))E(I_{31}(t,x)) = (\sqrt{n}/2)E[U_h(T,X,Y;T_1,X_1,Y_1;t,x)]$, we have

$$\begin{aligned} I_{31}(t,x) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} & E[U_h(T,X,Y;t_j,x_j,y_j;t,x)] \\ & -\frac{\sqrt{n}}{2}E[U_h(T,X,Y;T_1,X_1,Y_1;t,x)] + O_p(\frac{1}{\sqrt{nh}}). \end{aligned} \tag{A.4}$$

From the definition of U_h, condition (1) and some calculation, we have, letting $w_1(\cdot) = w(\cdot)f(\cdot)$,

$$\begin{aligned} E[U_h(T,X,Y;T_1,X_1,Y_1;t,x)] &= 2E[g(T_1)w_1(T)I(T\le t, X\le x)k_h(T_1-T)] \\ &= 2E[g(T+hu)w_1(T)I(T\le t, X\le x)K(u)] \\ &= 2E[g(T)w_1(T)I(T\le t, X\le x)] + O(h^2), \end{aligned} \tag{A.5}$$

$$\begin{aligned} & EU_h(T,X,Y;t_j,x_j,y_j;t,x) \\ = \ & w_1(t_j)I(t_j\le t, x_j\le x)\int g(hu+t_j)f(hu+t_j)K(u)d\,u \\ & +(y_j-\beta' x_j)\int F(x|T)f(T)w_1(T)I(T\le t)K((T-t_j)/h)/h\,dT \\ =: \ & a_j^{(1)}(t,x) + a_j^{(2)}(t,x), \end{aligned}$$

where $F(X|T)$ is the conditional distribution of X given T.

Define

$$\begin{aligned} b_j^{(1)}(t,x) &= g(t_j)f(t_j)w_1(t_j)I(t_j\le t, x_j\le x), \\ b_j^{(2)}(t,x) &= (y_j-\beta' x_j)F(x|t_j)f(t_j)w_1(t_j)I(t_j\le t). \end{aligned} \tag{A.6}$$

We have, from conditions (1)−(4) and (6),

$$\sup_{t,x} E(a_1^{(1)}(t,x) - b_1^{(1)}(t,x))^2 = O(h^2),$$
$$\sup_{t,x} E(a_1^{(2)}(t,x) - b_1^{(2)}(t,x))^2 = O(h^2). \tag{A.7}$$

Recalling $w_1(t) = w(t)f(t)$ we have

$$E(b_j^{(1)}(t,x)) = E(b_j^{(2)}(t,x)) = E[g(t)w(T)I(T \le t, X \le x)],$$
$$E(a_j^{(1)}(t,x) + a_j^{(2)}(t,x)) = E[U_h(T,X,Y;T_1,X_1,Y_1;t,x)]. \tag{A.8}$$

Let $c_j(t,x) = (a_j^{(1)}(t,x) + a_j^{(2)}(t,x) - b_j^{(1)}(t,x) - b_j^{(2)}(t,x))$. We now show that, uniformly over t and x,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (c_j(t,x) - Ec_j(t,x)) = O_p(h^{1/2}\log n + h^2\sqrt{n}). \tag{A.9}$$

It is easy to see that $\sup_{t,x} var(c_j(t,x)) \le \sup_{t,x} Ec_j^2(t,x) \le 2\sup_{t,x} E[a_1^{(1)}(t,x) - b_1^{(1)}(t,x))^2 + 2E(a_1^{(2)}(t,x) - b_1^{(2)}(t,x)]^2 \le O(h^2)$. Recall the class of all functions $c_j(t,x) = c(T_j,X_j,Y_j,t,x)$ with indices (t,x) discriminates finitely many points at a polynomial rate (that is, the class is a VC class), see Gaenssler (1983). The application of the symmetrization approach and the Hoeffding Inequality (see Pollard (1984, pp.14-16), yield that, for any $\delta > 0$ and some $w > 0$,

$$P\{\sup_{t,x} \frac{1}{\sqrt{n}} \sum_{j=1}^n (c_j(t,x) - Ec_j(t,x)) \ge \delta\}$$
$$\le 4E\{P\{\sup_{t,x} \frac{1}{\sqrt{n}} \sum_{j=1}^n \sigma_j(c_j(t,x) - Ec_j(t,x)) \ge \delta/4|T_j,X_j,Y_j, j=1,\ldots,n\}\}$$
$$\le E\{(cn^w \sup_{t,x} exp([-\frac{\delta^2}{32\frac{1}{n}\sum_{j=1}^n (c_j(t,x) - Ec_j(t,x))^2}]) \wedge 1\}.$$

In order to prove the above to be asymptotically zero, we now bound the denominator in the power. Applying condition (1) and the Uniformly Strong Law of Large Number, see Pollard (1984, p.23, Chap. II Th. 24), $\sup_{t,x}(1/n)\sum_{j=1}^n (c_j(t,x) - Ec_j(t,x))^2 = O_p(h)$. Letting $\delta = h^{1/2}\log n$ we can derive (A.9). Furthermore by (A.7), it is easy to verify that $\sqrt{n}E[c_i(t,x)] = O(\sqrt{n}h^2) = o(1)$. Combining with (A.6) and (A.9), we see that

$$I_{31}(t,x) = \frac{1}{\sqrt{n}} \sum_{j=1}^n (b_j^{(1)}(t,x) + b_j^{(2)}(t,x) - Eg(T)w(T)I(T \le t, X \le x))$$

$$+O_p(\frac{1}{\sqrt{nh}}+h)$$
$$= I_4(t,x)+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(b_j^{(2)}(t,x)-Eg(T)w(T)I(T\le t, X\le x))$$
$$+O_p(\frac{1}{\sqrt{nh}}+h). \tag{A.10}$$

We now turn to I_{32}. Following the above U-statistic argument, and using (1) ad $w_2(\cdot)=w(\cdot)g(\cdot)$ in the lieu of $y_j-\beta' x_j$ and $w(\cdot)$, respectively, we can verify that

$$I_{32}(t,x)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(g(t_j)w(t_j)(I(t_j\le t, x_j\le x)$$
$$+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(g(t_j)w(t_j)F(x|t_j)T(t_j\le t))-E[g(T)w(T)I(T\le t, X\le x)]\Big)$$
$$+O_p(\frac{1}{\sqrt{nh}}+h)$$
$$= I_{33}(t,x)+\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(g(t_j)w(t_j)F(x|t_j)I(t_j\le t)$$
$$-E[g(T)w(T)I(T\le t, X\le x)]\Big)$$
$$+O_p(\frac{1}{\sqrt{nh}}+h). \tag{A.11}$$

By (A.2), (A.10) and (A.11),

$$I_3(t,x)-I_4(t,x)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon w(t_j)F(x|t_j)I(t_j\le t)+O_p(\frac{1}{\sqrt{nh}}+h). \tag{A.12}$$

This clearly converges in distribution to a Gaussian process. The proof is concluded from (A.2) and (A.12).

Proof of Theorem 2.2. Following the lines of Schick (1996), it is easy to see that $\sqrt{n}(\hat\beta-\beta)=S^{-1}\{(1/\sqrt{n})\sum_{j=1}^{n}U(t_j,x_j)\varepsilon_j w^2(t_j)\}+C_1+o_p(1)$, where $C_1=S^{-1}E[U(T,X)\ (g_1(T,X)-E(g_1(T,X)|T))w^2(T)]$. Similar to the proof of Theorem 2.1, it follows that

$$R_n(t,x)\ =\ \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j w(t_j)I(t_j\le t, x_j\le x)$$
$$-E(U(T,X)'w(T)I(T\le t, X\le x)S^{-1}[\frac{1}{\sqrt{n}}\sum_{j=1}^{n}U(t_j,x_j)\varepsilon_j w^2(t_j)+C]$$

$$-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\hat{g}(t_j)-g(t_j))w(t_j)I(t_j\le t,x_j\le x)+O_p(\frac{1}{h\sqrt{n}}+\sqrt{n}h^2)$$
$$+\frac{1}{n}\sum_{j=1}^{n}(g_1(t_j,x_j)-E(g_1(T,X)|T=t_j))w(t_j)I(t_j\le t,x_j\le x)$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j w(t_j)I(t_j\le t,x_j\le x)$$
$$-E(U(T,X)'w(T)I(T\le t,X\le x)S^{-1}[\frac{1}{\sqrt{n}}\sum_{j=1}^{n}U(t_j,x_j)\varepsilon_j w^2(t_j)]$$
$$-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\hat{g}(t_j)-g(t_j))w(t_j)I(t_j\le t,x_j\le x)$$
$$+g_{1*}(t,x)+O_p(\frac{1}{h\sqrt{n}}+\sqrt{n}h^2)$$
$$=: J_1(t,x)-J_2(t,x)-J_3(t,x)+g_{1*}(t,x)+O_p(\frac{1}{h\sqrt{n}}+\sqrt{n}h^2),$$

where $g_{1*}(t,x)$ is defined in Theorem 2.2. Let $\tilde{Y}=\beta'X+g(T)+\varepsilon$, $g_2(t)=E(\tilde{Y}-\beta'X|T=t)$ and, pretending $\tilde{Y}$ is observable, its estimator at point t_i, $\hat{g}_2(t_i)=\hat{E}_i(\tilde{Y}-\beta'X|T=t_i)$, similar to that in (2.6). It is clear that $g(t_i)=g_2(t_i)+E(g_1(T,X)|T=t_i)/\sqrt{n}$ and $\hat{g}(t_i)=\hat{g}_2(t_i)+\hat{E}_i(g_1(T,X)|T=t_i)/\sqrt{n}$. Furthermore, $\sup_i \hat{E}_i(g_1(T,X)|T=t_i)-E(g_1(T,X)|T=t_i)|\to 0$ in probability as $n\to\infty$. Then

$$J_3(t,x)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\hat{g}_2(t_j)-g_2(t_j))w(t_j)I(t_j\le t,x_j\le x)$$
$$+\frac{1}{n}\sum_{j=1}^{n}[\hat{E}_j(g_1(T,X)|T=t_j)-E(g_1(T,X)|T=t_j)]w(t_j)I(t_j\le t,x_j\le x)$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}[\hat{g}_2(t_j)-g_2(t_j)]w(t_j)I(t_j\le t,x_j\le x)+o_p(1).$$

Therefore J_3 is asymptotically equal to $I_3(t,x)-I_4(t,x)$ and J_1 and J_2 are analogous to I_1 and I_2 in (A.2). The proof follows from the argument for proving Theorem 2.1.

A.3. Proof, Section 3

Proof of Theorem 3.1. We need only notice that even under the local alternative, $\hat{\beta}$, $\hat{U}$, $\hat{S}$ and $\hat{F}(X|T)$ are consistent for β, U, S and $F(X|T)$.

By Wald's device, for almost all sequences $\{(t_1,x_1,y_1),\ldots,(t_n,x_n,y_n),\ldots\}$, we need to show that (i) the covariance function of $R_n(E_n)$ converges to that of R, (ii) finite distributional convergence of $R_n(E_n)$ holds for any finite indices $(t_1, x_1), \ldots, (t_k, x_k)$ and (iii) uniform tightness. The properties (i) and (ii) are easily verified, the details are omitted. For (iii) we notice that the functions $J(\cdot, t, x)$ over all indices (t, x) is a VC class of functions. The Equicontinuity Lemma holds, see Pollard (1984, p.150). By Theorem VII 21 of Pollard (1984, p.157), $R_n(E_n)$ converges in distribution to a Gaussian process R. For more details, a similar augument can be found in Zhu, Fujikoshi and Naito (2001, the proof of Theorem 3.2). The proof is finished.

Acknowledgement

This research was supported by the NNSF of China, a UGC grant of University of Hong Kong and a RGC grant of Hong Kong, HKSAR, China. We are grateful to the Editor, an associate editor and two referees for their suggestions which have led to a great improvement in the paper. We also thank a referee for providing us a reference, and Wolfgang Härdle for his constructive comments when he visited the University of Hong Kong.

References

Cuzick, J. (1992). Semi-parametric additive regression. *J. Roy. Statist. Soc. Ser. B* **54**, 831-843.

Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.* **27**, 1012-1040.

Dudley, R. M. (1978). Central limit theorems for empirical measures. *Ann. Probab.* **6**, 899-929.

Engle, R. F., Granger, C. W. J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **81**, 310-320.

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

Fan, J. and Huang, L. (2001). Goodness-of-fit tests for parametric regression models. *J. Amer. Statist. Assoc.* **96**, 310-320.

Fan, Y. and Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica* **64**, 865-890.

Gozalo, P. L. and Linton, O. B. (2001). Testing additivity in generalized nonparametric regression models with estimated parameters. *J. Econometrics* **104**, 1-48.

Gaensler, P. (1983). Empirical processes. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 3. Institute of Mathematical Statistics, Hayward, California, USA.

Gine, E. and Zinn, J. (1984). On the central limit theorem for empirical processes (with discussion). *Ann. Probab.* **12**, 929-998.

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

Härdle, W. and Mammen, E. (1993). Comparing non-parametric versus parametric regression fits. *Ann. Statist.* **21**, 1926-1947.

Mammen, E. and van de Geer, S. (1997). Penalized quasi-likelihood estimation in partial linear models. *Ann. Statist.* **25**, 1014-1035.

Neuhaus, G. and Zhu, L. X. (1998). Permutation test for symmetry. *J. Multivariate Anal.* **67**, 129-153.

Nolan, D. and Pollard, D. (1987). U-process: Rates of convergence. *Ann. Statist.* **15**, 780-799.

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

Schick, A. (1996). Root-n consistent estimation in partly linear regression models. *Statist. Probab. Lett.* **28**, 353-358.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. Ser. B* **50**, 413-436.

Spoikoiny, V. G. (1996). Adaptive hypotheesis testing using wavelets. *Ann. Statist.* **24**, 2477-2498.

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25**, 613-641.

Stute, W., Manteiga, G. and Quindimil, M. P. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93**, 141-149.

Stute, W., Thies, S. and Zhu, L. X. (1998). Model checks for regression: An innovation process approach. *Ann. Statist.* **26**, 1916-1934.

Sykes, L. R., Isacks, B. L. and Oliver, J. (1969). Spatial distribution of deep and shallow earthquakes of small magnitudes in the Fiji-Tonga region. *Bull. Seismol. Soc. Amer.* **59**, 1093-1113.

van der Vaart, A. W. and Wellner, J. A. (2000). *Weak Convergence and Empirical Processes*. Springer, New York.

Whang, Y. and Andrews, D. W. K. (1993). Tests of specification for parametric and semiparametric models. *J. Econometrics* **57**, 277-318.

Yatchew, A. J. (1992). Nonparametric regression tests based on least squares. *Econom. Theory* **8**, 435-451.

Zhu, L. X. and Ng, K. W. (2002). Checking the adequacy of a partially linear model. Tech. Report # 342, Department of Statistics and Actuarial Science, The University of Hong Kong.

Zhu, L. X. and Fang, K. T. (1996). Asymptotics for kernel estimate of sliced inverse regression. *Ann. Statist.* **14**, 1053-1068.

Zhu, L. X., Fujikoshi, Y. and Naito, K. (2001). Heteroscedasticity test for regression models. *Sci. China Ser. A* **44**, 1237-1252.

Zhu, L. X., Ng, W. and Jing, P. (2002). Resampling methods for homogeneity tests of covariance matrices. *Statist. Sinica* **12**, 769-783.

Zhu, L. X. (2003). Model checking of dimension-reduction type for regression. *Statist. Sinica* **13**, 283-296.

Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, 100080, China.

E-mail: lzhu@hku.hk

Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong.

E-mail: kaing@hku.hk

(Received January 2002; accepted April 2003)