Statistica Sinica **24** (2014), 937-955
doi:http://dx.doi.org/10.5705/ss.2012.127

SIMULTANEOUS CONFIDENCE BANDS AND HYPOTHESIS TESTING FOR SINGLE-INDEX MODELS

Gaorong Li[1], Heng Peng[2], Kai Dong[2] and Tiejun Tong[2]

[1]*Beijing University of Technology and* [2]*Hong Kong Baptist University*

Abstract: In this paper, we propose simultaneous confidence bands for the nonparametric link function in single-index models in the presence of a nuisance index parameter. We establish the asymptotic properties for the link function and its derivative that allow simultaneous confidence bands for various inference tasks. In addition, we propose an adaptive Neyman test statistic for testing the linearity of the link function. We then conduct simulation studies to evaluate the performance of the proposed method, and apply them to two data sets for illustration.

Key words and phrases: Adaptive Neyman test, difference-based estimator, local linear smoother, residual variance, simultaneous confidence band, single-index model.

1. Introduction

In essence, a single-index model (SIM) is a linear regression model with a nonparametric link function. It provides a parsimonious way to implement multivariate nonparametric regression, and avoids the so-called "curse of dimensionality". In this paper, we consider the SIM

$$Y = \eta(\boldsymbol{\beta}_0^T \boldsymbol{X}) + \varepsilon, \tag{1.1}$$

where $\eta(\cdot)$ is an unknown link function, $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ is a $p \times 1$ unknown index vector, $\boldsymbol{X}$ is a p-dimensional covariate vector, and ε is a random error with mean zero and variance σ^2. We assume that $\boldsymbol{X}$ and ε are independent of each other, and $\|\boldsymbol{\beta}_0\| = 1$ with the first element of $\boldsymbol{\beta}_0$ positive to ensure identifiability. Single-index models (SIMs) are popular and efficient in modeling high-dimensional data, and there is a large literature on this topic. See, for example, Härdle, Hall, and Ichimura (1993), Carroll et al. (1997), Xia and Li (1999), Xia et al. (2002), Xia et al. (2004), Yu and Ruppert (2002), Zhu and Xue (2006), Wang et al. (2010), Li et al. (2010), Liang et al. (2010), Chen, Gao, and Li (2013), Cui, Härdle, and Zhu (2011), among others.

A problem in nonparametric or semiparametric regression is the construction of simultaneous confidence bands (SCBs) for the mean function.

Bickel and Rosenblatt (1973) considered SCBs and a goodness-of-fit test for the estimation of a density function. Härdle and Bowman (1988) considered SCBs in combination with locally adaptive smoothing parameters. Sun and Loader (1994) constructed conservative confidence bands for the mean function based on some prior information about the maximum roughness of mean function. Faraway and Sun (1995) considered SCBs for heteroscedastic models using kernel smoothing. Fan and Zhang (2000) and Zhang and Peng (2010) considered SCBs for varying-coefficient models with a plug-in estimate of bias. Zhang, Fan, and Sun (2009) studied the semiparametric model with cluster data by accounting for within-cluster correlation. Krivobokova, Kneib, and Claeskens (2010) constructed SCBs for univariate smooth curves based on penalized spline estimators. Further reference in constructing SCBs with heteroscedastic errors can be seen in Neumann and Polzehl (1998), Claeskens and Van Keilegom (2003), and more.

For most nonparametric and semiparametric models, the nonparametric component estimation can be simplified to a nonparametric regression estimation with estimates of the nonparametric components orthogonal to estimates of the parametric components. Therefore, the classical statistical inference methods for nonparametric regression models can be readily extended to general nonparametric and semiparametric models.

For SIMs, however, we do not have the orthogonality property between the estimated link function and the estimated index parameter. To our knowledge, there is little work in the literature in constructing SCBs for SIMs when the index vector is treated as a nuisance parameter. SCBs for SIMs are useful for checking the graphical representation of the link function. To improve the efficiency for SCBs, we take into account the bias and high-order derivatives of the link function estimates. The bias-corrected technique was introduced by Eubank and Speckman (1993) for local constant kernel estimation, and by Xia (1998) for local linear estimation in nonparametric regression.

The paper is organized as follows. In Section 2, we propose the estimation procedures for model (1.1), and establish the asymptotic properties of the proposed estimator. Specifically, we estimate the link function by the local linear smoother, and the index parameter by the profile least squares method. In Section 3, we test the hypothesis that the link function is linear. To achieve this, we propose to extend the adaptive Neyman test of Fan (1996) to SIMs. We then conduct simulation studies in Section 4 to evaluate the performance of the proposed method, and apply them to two data sets in Section 5 for illustration. We conclude the paper in Section 6 with some remarks, and present the technical details in the supplementary material.

2. Estimation Procedure

2.1. Estimation of the link function

Consider the SIM,

$$Y_i = \eta(\boldsymbol{\beta}_0^T \boldsymbol{X}_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where Y_i are observations, $\boldsymbol{X}_i$ are covariate vectors, and ε_i are independent and identically distributed random errors with mean zero and variance σ^2. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$.

Of the various methods to estimate the link function η and its derivative η', we apply the local linear regression technique proposed in Fan and Gijbels (1996). It is known that the local linear fitting has advantages in such aspects as high asymptotic efficiency, design adaption, and automatic boundary correction. Let $K(\cdot)$ be a kernel function and $K_h(\cdot) = h^{-1}K(\cdot/h)$, where $h = h_n$ is the bandwidth. For a given $\boldsymbol{\beta}$, the estimators of η and η' are obtained by minimizing the weighted sum of squares

$$\sum_{i=1}^{n} \{Y_i - a - b(\boldsymbol{\beta}^T \boldsymbol{X}_i - u)\}^2 K_h(\boldsymbol{\beta}^T \boldsymbol{X}_i - u) \tag{2.2}$$

with respect to a and b. Specifically, the local linear estimators of $\eta(u)$ and $\eta'(u)$ are $\hat{\eta}(u;\boldsymbol{\beta}) = \hat{a}$ and $\hat{\eta}'(u;\boldsymbol{\beta}) = \hat{b}$ for the given $\boldsymbol{\beta}$. Let $\boldsymbol{\eta}(u) = (\eta(u), h\eta'(u))^T$, $\hat{\boldsymbol{\eta}}(u) = (\hat{\eta}(u;\boldsymbol{\beta}), h\hat{\eta}'(u;\boldsymbol{\beta}))^T$, and

$$\boldsymbol{W} = \mathrm{diag}\Big(K_h(\boldsymbol{\beta}^T \boldsymbol{X}_1 - u), \ldots, K_h(\boldsymbol{\beta}^T \boldsymbol{X}_n - u)\Big), \quad \mathbf{X}_h = \begin{pmatrix} 1 & \frac{\boldsymbol{\beta}^T \boldsymbol{X}_1 - u}{h} \\ \vdots & \vdots \\ 1 & \frac{\boldsymbol{\beta}^T \boldsymbol{X}_n - u}{h} \end{pmatrix}.$$

By least squares, we have

$$\hat{\boldsymbol{\eta}}(u) = (\mathbf{X}_h^T \boldsymbol{W} \mathbf{X}_h)^{-1} \mathbf{X}_h^T \boldsymbol{W} \boldsymbol{Y} = \boldsymbol{S}_n(u;\boldsymbol{\beta})^{-1} \boldsymbol{V}_n(u;\boldsymbol{\beta}), \tag{2.3}$$

where $\boldsymbol{S}_n(u;\boldsymbol{\beta}) = \begin{pmatrix} S_{n,0}^{\boldsymbol{\beta}} & S_{n,1}^{\boldsymbol{\beta}} \\ S_{n,1}^{\boldsymbol{\beta}} & S_{n,2}^{\boldsymbol{\beta}} \end{pmatrix}$ and $\boldsymbol{V}_n(u;\boldsymbol{\beta}) = (V_{n,0}^{\boldsymbol{\beta}}, V_{n,1}^{\boldsymbol{\beta}})^T$ with

$$S_{n,l}^{\boldsymbol{\beta}} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\boldsymbol{\beta}^T \boldsymbol{X}_i - u}{h} \right)^l K_h(\boldsymbol{\beta}^T \boldsymbol{X}_i - u), \qquad l = 0, 1, 2, \tag{2.4}$$

$$V_{n,l}^{\boldsymbol{\beta}} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\boldsymbol{\beta}^T \boldsymbol{X}_i - u}{h} \right)^l K_h(\boldsymbol{\beta}^T \boldsymbol{X}_i - u) Y_i, \qquad l = 0, 1. \tag{2.5}$$

With some straightforward algebra, we have

$$\hat{\boldsymbol{\eta}}(u) = \boldsymbol{\eta}(u) + (\mathbf{X}_h^T \boldsymbol{W} \mathbf{X}_h)^{-1}(\frac{A_n}{2} + C_n) + o_P\Big(\frac{1}{\sqrt{nh}}\Big),$$

where $A_n = \mathbf{X}_h^T \boldsymbol{W} \boldsymbol{r}(u,h) \cdot h^2 \eta''(u)$, $C_n = \mathbf{X}_h^T \boldsymbol{W} \boldsymbol{\varepsilon}$, and

$$\boldsymbol{r}(u,h) = \bigg(\Big(\frac{\boldsymbol{\beta}^T \boldsymbol{X}_1 - u}{h}\Big)^2, \ldots, \Big(\frac{\boldsymbol{\beta}^T \boldsymbol{X}_n - u}{h}\Big)^2\bigg)^T.$$

To construct SCBs for η and η', we need $\sqrt{n}-$consistent estimators of the index vector $\boldsymbol{\beta}_0$ and the variance σ^2. There are different methods for estimating $\boldsymbol{\beta}_0$ consistently. See, for example, Härdle and Stoker (1989), Härdle, Hall, and Ichimura (1993), Weisberg and Welsh (1994), Carroll et al. (1997), and Wang et al. (2010). We employ the profile least squares method (Liang et al. (2010); Wang et al. (2010)) and that results in

$$\hat{\boldsymbol{\beta}}_0 = \underset{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|=1}{\operatorname{argmin}} \sum_{i=1}^n \{Y_i - \hat{\eta}(\boldsymbol{\beta}^T \boldsymbol{X}_i; \boldsymbol{\beta})\}^2. \tag{2.6}$$

The Newton-Raphson iterative method can be used to update the estimators of $\boldsymbol{\beta}_0$ and η. In addition, the efficiency of the estimation can be improved by using the re-parametrization technique under the constraint $\|\boldsymbol{\beta}_0\| = 1$. The resulting estimator is $\sqrt{n}-$consistent and asymptotically normal (Liang et al. (2010); Wang et al. (2010); Cui, Härdle, and Zhu (2011)). Finally, with the plug-in estimator $\hat{\boldsymbol{\beta}}_0$ in (2.3), we have the estimator

$$\hat{\boldsymbol{\eta}}(u; \hat{\boldsymbol{\beta}}_0) = (\hat{\eta}(u; \hat{\boldsymbol{\beta}}_0),\ h\hat{\eta}'(u; \hat{\boldsymbol{\beta}}_0))^T = \boldsymbol{S}_n(u; \hat{\boldsymbol{\beta}}_0)^{-1} \boldsymbol{V}_n(u; \hat{\boldsymbol{\beta}}_0). \tag{2.7}$$

2.2. Asymptotic properties

Let $\delta_n = \sqrt{\log n/(nh)}$, $\mu_l = \int t^l K(t)dt$, and $\nu_l = \int t^l K^2(t)dt$ for $l = 0, 1, 2$. Let $\|g\|_\infty = \sup_{u \in [b_1, b_2]} |g(u)|$ for any function $g(u)$, and $\|A\|_\infty = (\sum_{i=1}^p \sum_{j=1}^p \|a_{ij}\|_\infty^2)^{1/2}$ for any square matrix $A(u) = (a_{ij}(u))$ of size p. To achieve the asymptotic results, we need the following regularity conditions.

C1. $\{\boldsymbol{X}_i, \varepsilon_i\}, i \geq 1$, are independent, and $E(\varepsilon) = 0, \mathrm{Var}(\varepsilon) = \sigma^2 < \infty, E(\varepsilon^4) < \infty$ and $E\|\boldsymbol{X}_i\|^{2+\tau} < \infty$ for some $\tau > 0$.

C2. The function $\eta(\cdot)$ is differentiable on the compact set $\mathcal{U} = [b_1, b_2]$ of $\boldsymbol{\beta}^T \boldsymbol{X}$.

C3. The function $\eta(\boldsymbol{\beta}^T \boldsymbol{X})$ and the density function of $\boldsymbol{\beta}^T \boldsymbol{X}$, $f(u)$, are three times continuously differential with respect to u. The third derivatives are uniformly Lipschitz continuous over $\mathcal{A} \subset \mathbb{R}^p$ for all $u \in \{u = \boldsymbol{\beta}^T \boldsymbol{x} : \boldsymbol{\beta} \in \mathcal{A}, \boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^p\}$.

C4. The kernel function $K(\cdot)$ is a symmetric density function, absolutely continuous on its support set $[-c_0, c_0]$.

C4a. $K(c_0) \neq 0$ or

C4b. $K(c_0) = 0, K(z)$ is absolutely continuous and $K^2(z)$, $(K'(z))^2$ are integrable on $(-\infty, \infty)$.

C5. The bandwidth h satisfies $nh^3/\log n \to \infty$ and

C5a. $nh^5 \log n \to 0$ as $n \to \infty$;

C5b. $nh^7 \log n \to 0$ as $n \to \infty$.

C6. There exists an integer n_0 such that

$$\left\{\frac{n}{(\log\log n)^4}\right\}^{-1} \sum_{i=n_0}^{n/(\log\log n)^4} X_{ij}^{*2} = O_P(1), \quad j = 1, \ldots, p,$$

where X_{ij}^* is the jth element of the vector $\boldsymbol{X}_i^*$, and $\boldsymbol{X}_i^*$ is the ith row of the $n \times p$ matrix $\mathbf{X}^* = \Gamma\mathbf{X}$.

C2 and C3 are standard smoothness conditions. C4 is a set of mild conditions on the kernel function that have been used by many authors (see Fan and Zhang (2000); Zhang, Fan, and Sun (2009)). The bandwidth condition C5a and C5b were used in Claeskens and Van Keilegom (2003). C6 is a mild condition that can be found in Fan and Huang (2001); it holds almost surely for designs generated from a random sample.

Theorem 1. *Under* C1–C5a*, for all $u \in [b_1, b_2]$ we have*

$$\|\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0\| = O_P(n^{-1/2}), \tag{2.8}$$

$$\sqrt{nh}\Big\{\hat{\eta}(u; \hat{\boldsymbol{\beta}}_0) - \eta(u) - b(u)\Big\} \xrightarrow{D} N\Big(0, \frac{\nu_0}{f(u)}\sigma^2\Big), \tag{2.9}$$

where $\xrightarrow{D}$ denotes the convergence in distribution, and $b(u) = h^2\mu_2\eta''(u)/2$.

Theorem 2. *Suppose that $h = O(n^{-\rho})$ with $1/5 < \rho < 1/3$ for all $u \in [b_1, b_2]$. If $B = (-2\log\{h/(b_2 - b_1)\})^{1/2}$, under* C1–C5a*, we have*

$$P\Big\{B\Big((\sigma^2\nu_0)^{-1/2} \sup_{u\in[b_1,b_2]} \Big|(nhf(u))^{1/2}\big(\hat{\eta}(u; \hat{\boldsymbol{\beta}}_0) - \eta(u) - b(u)\big)\Big| - d_{n0}\Big) < x\Big\}$$
$$\longrightarrow \exp\Big(-2\exp(-x)\Big), \qquad \text{as } n \to \infty,$$

where

$$d_{n0} = \begin{cases} B + B^{-1}\Big\{\log\frac{K^2(c_0)}{\nu_0\pi^{1/2}} + \frac{1}{2}\log\log\left(\frac{b_2-b_1}{h}\right)\Big\}, & \text{if } K(c_0) \neq 0, \\ B + B^{-1}\log\left\{\frac{1}{4\nu_0\pi}\int(K'(t))^2dt\right\}, & \text{if } K(c_0) = 0. \end{cases}$$

Theorem 2 gives the asymptotic distribution of the maximum absolute deviation between the estimated link function and the true link function when the estimator of the index parameter is $\sqrt{n}$−consistent. The convergence rate of $\hat{\eta}'(\cdot)$ can be relatively slow, is usually sensitive to the choice of bandwidth, and thus requires a larger weight on the support set. Therefore, we only present the asymptotic distribution of the maximum absolute deviation for $\hat{\eta}'(\cdot)$ under C4b.

Theorem 3. *Under* C1−C5a*, for all* $u \in [b_1, b_2]$ *we have*

$$P\Big\{B\Big((\sigma^2\nu_2)^{-1/2}\sup_{u\in[b_1,b_2]}\Big|(nh^3 f(u)\mu_2^2)^{1/2}\Big(\hat{\eta}'(u;\hat{\boldsymbol{\beta}}_0)-\eta'(u)\Big)\Big|-d_{n1}\Big)<x\Big\}$$
$$\longrightarrow \exp\big(-2\exp(-x)\big), \quad \text{as} \quad n\to\infty,$$

where B *is defined in Theorem* 2 *and if* $K(c_0)=0$*,*

$$d_{n1}=B+B^{-1}\log\Big\{\frac{1}{2\pi\sqrt{\nu_2}}\Big(\int t^2(K'(t))^2dt\Big)^{1/2}\Big\}.$$

2.3. Estimation of bias and variance

To derive the bias and variance of $\hat{\eta}$, we need the estimates of the unknown $f(u)$ and σ^2. For $f(u)$, we propose

$$\hat{f}(u)=\frac{1}{nh}\sum_{i=1}^{n}K\Big(\frac{\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i-u}{h}\Big),$$

where $\hat{\boldsymbol{\beta}}_0$ is the profile least squares estimator defined by (2.6).

We employ the least squares estimator in Tong and Wang (2005) to estimate σ^2 without estimating the mean function in the model. For single-index models there are many methods to efficiently estimate $\boldsymbol{\beta}_0$ without estimating the link function $\eta(\cdot)$. As shown in Section 3, if σ^2 can be estimated efficiently without estimating the link function $\eta(\cdot)$, our adaptive Neyman test for the link function $\eta(\cdot)$ avoids estimating $\eta(\cdot)$ directly. Hence the proposed method is computationally efficient and the results are stable.

Let $\hat{\boldsymbol{\beta}}_0$ be a consistent and efficient estimate of $\boldsymbol{\beta}_0$. We reorder Y_i as Y_i^* according to the monotonic order of $\boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}_0, i=1,\ldots,n$, and let $s_k=\sum_{i=k+1}^{n}(Y_i^*-Y_{i-k}^*)^2/\{2(n-k)\}$ be the average of $(n-k)$ lag-k differences, where $1\le k\le m$. Tong and Wang (2005) observed that $E(s_k)=\sigma^2+J\zeta_k+o(\zeta_k)$ where $J=\int_0^1\{\eta'(x)\}^2dx/2$ and $\zeta_k=k^2/n^2$. They proposed to regress s_k on ζ_k by a simple linear model and then to estimate σ^2 as the intercept. The fitted intercept was shown to be a consistent and optimally efficient estimator of σ^2 as long as $m\to\infty$ and $m/n\to 0$. While m serves as a tuning parameter and the optimal value can be selected by the data-driven $V-$fold cross validation, when $m = n^{\frac{1}{2}}$ such least squares estimator of σ^2 performs well as long as the link function or mean function in the model is smooth enough, see Section 3.1 in Tong and Wang (2005) for more details.

Theorem 1 gives the asymptotic bias and variance of $\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)$ and, with the consistent estimators $\hat{f}(u)$ and $\hat{\sigma}^2$, we estimate the variance of $\hat{\eta}$ by $\widehat{\mathrm{Var}}\{\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)|\mathcal{D}\} = \nu_0\hat{\sigma}^2/[nh\hat{f}(u)]$, where $\mathcal{D} = (\boldsymbol{X}_1^T,\ldots,\boldsymbol{X}_n^T)^T$. We estimate the bias of $\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)$ by

$$\widehat{\mathrm{bias}}(\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)|\mathcal{D}) = \frac{h^2\mu_2\hat{\eta}''(u;\hat{\boldsymbol{\beta}}_0)}{2}, \tag{2.10}$$

where the estimate $\hat{\eta}''(u;\hat{\boldsymbol{\beta}}_0)$ is obtained by using the local cubic fit with an appropriate pilot bandwidth. We choose $h_* = O(n^{-1/7})$ since it is optimal for estimating $\eta''(u)$ and it can be chosen by the residual squares criterion (Fan and Gijbels (1996)).

2.4. SCBs for the link function

Theorem 4. *Suppose that $h = O(n^{-\rho})$ with $1/5 < \rho < 1/3$ for all $u \in [b_1,b_2]$, $\eta^{(3)}(\cdot)$ is continuous on $[b_1,b_2]$, and the pilot bandwidth h_* is of order $n^{-1/7}$. Under* C1$-$C5a*, for all $u \in [b_1,b_2]$ we have*

$$P\Big\{B\Big(\sup_{u\in[b_1,b_2]}\Big|\frac{\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)-\eta(u)-\widehat{\mathrm{bias}}(\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)|\mathcal{D})}{[\widehat{\mathrm{Var}}\{\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)|\mathcal{D}\}]^{1/2}}\Big|-d_{n0}\Big)<x\Big\}$$
$$\to \exp\big(-2\exp(-x)\big), \qquad \text{as } n\to\infty,$$

where B and d_{n0} are as defined in Theorem 2.

By Theorem 4, we have a $100(1-\alpha)\%$ confidence band for $\eta(u)$,

$$\Big(\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0) - \widehat{\mathrm{bias}}(\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)|\mathcal{D}) \pm \Delta_{1,\alpha}(u)\Big), \tag{2.11}$$

where $\Delta_{1,\alpha}(u) = (d_{n0} + B[\log 2 - \log\{-\log(1-\alpha)\}])[\widehat{\mathrm{Var}}\{\hat{\eta}(u;\hat{\boldsymbol{\beta}}_0)|\mathcal{D}\}]^{1/2}$.

An application of Theorem 4 addresses the graphical questions about the link function. Thus, if the $100(1-\alpha)\%$ SCB for $\eta(u)$ over the set $[b_1,b_2]$ does not contain a linear function, we conclude that the link function $\eta(\cdot)$ is nonlinear. An appropriate choice of $[b_1,b_2]$ is important to the performance of the proposed method for constructing the SCB of the link function. By Theorems 2$-$4, such interval can be treated as the compact support of $\boldsymbol{X}_i^T\boldsymbol{\beta}_0$. In practice when $\boldsymbol{\beta}_0$ is unknown, we can take $b_1 = \min\{\boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}_0, i = 1,\ldots,n\}$ and $b_2 = \max\{\boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}_0, i = 1,\ldots,n\}$ if the link function is smooth and the boundary effect of the nonparametric estimation for the link function is negligible. Otherwise, to remove the boundary effect, b_1 and b_2 can be set as 5% to 10 % and 90% to 95% quantiles of $\{\boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}_0, i = 1, \ldots, n\}$, respectively. In our numerical study in Section 4, the estimated link function is quite smooth, and the boundary effect is automatically corrected when the local polynomial regression is used to estimate the link function. We thus take b_1 and b_2 to be the minimum and maximum of $\{\boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}_0, i = 1, \ldots, n\}$ in our simulation studies.

3. The Adaptive Neyman Test

Consider the testing problem

$$H_0 : \eta(u) = \gamma_0 + \gamma_1 u \longleftrightarrow H_1 : \eta(u) \neq \gamma_0 + \gamma_1 u, \tag{3.1}$$

where γ_0 and γ_1 are two unknown parameters. Under H_0, $\eta'(u)$ is a constant and the test is equivalent to testing

$$H_0 : \eta'(u) = \gamma_1 \longleftrightarrow H_1 : \eta'(u) \neq \gamma_1. \tag{3.2}$$

By Theorem 3, a natural test statistic is

$$\Big(-2\log\Big\{\frac{h}{b_2 - b_1}\Big\}\Big)^{1/2}\Big\{\sup_{u\in[b_1,b_2]}\Big|\frac{\hat{\eta}'(u;\hat{\boldsymbol{\beta}}_0) - \hat{\gamma}_1}{[\widehat{\mathrm{Var}}\{\hat{\eta}'(u;\hat{\boldsymbol{\beta}}_0)|\mathcal{D}\}]^{1/2}}\Big| - d_{n1}\Big\}. \tag{3.3}$$

We reject H_0 when the test statistic exceeds the asymptotic critical value $c_\alpha = -\log\{-0.5\log(1-\alpha)\}$. The estimators of γ_0 and γ_1 can be obtained by the two-stage estimation procedure in Fan and Zhang (2000). We treat γ_0 and γ_1 as functions. For a given $\boldsymbol{\beta}$, we can obtain the estimators $\hat{\gamma}_0(\boldsymbol{\beta}^T\boldsymbol{X}_i;\boldsymbol{\beta})$ and $\hat{\gamma}_1(\boldsymbol{\beta}^T\boldsymbol{X}_i;\boldsymbol{\beta})$ for γ_0 and γ_1 at $\boldsymbol{\beta}^T\boldsymbol{X}_i(i = 1, \ldots, n)$ using weighted as at least squares (2.2). Updating the estimator of $\boldsymbol{\beta}_0$ by the profile least squares method (2.6) leads to estimators $\hat{\gamma}_0(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i;\hat{\boldsymbol{\beta}}_0)$ and $\hat{\gamma}_1(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i;\hat{\boldsymbol{\beta}}_0)$ over $i = 1, \ldots, n$. We take the average of $\hat{\gamma}_0(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i;\hat{\boldsymbol{\beta}}_0)$ and $\hat{\gamma}_1(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i;\hat{\boldsymbol{\beta}}_0)$ over $i = 1, \ldots, n$ to get

$$\hat{\gamma}_0 = \frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_0(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i;\hat{\boldsymbol{\beta}}_0), \quad \hat{\gamma}_1 = \frac{1}{n}\sum_{i=1}^{n}\hat{\gamma}_1(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i;\hat{\boldsymbol{\beta}}_0). \tag{3.4}$$

This two-steps estimation uses a smaller bandwidth in the first step to control the bias, and then averages in the second step to reduce the variance. The test statistic (3.3) is, however, difficult to implement due to the sensitive choice of an appropriate bandwidth.

Here is an alternative method for the test problem (3.1). Let $(\widetilde{\gamma}_0, \widetilde{\gamma}_1)$ be the least squares estimators, $\hat{Y}_i = \widetilde{\gamma}_0 + \widetilde{\gamma}_1(\hat{\boldsymbol{\beta}}_0^T\boldsymbol{X}_i)$ be the fitted value under H_0, with residuals $\hat{\varepsilon}_i = Y_i - \hat{Y}_i, i = 1, \ldots, n$. To improve power, Fan and Huang

(2001) suggested using the discrete Fourier transform of the residual vector $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)^T$ to compress useful signals into low frequencies. Let Γ be the $n \times n$ orthonormal matrix generated by the discrete Fourier transform, and $\hat{\boldsymbol{\varepsilon}}^* = \Gamma\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1^*, \ldots, \hat{\varepsilon}_n^*)^T$ be the residual vector through the discrete Fourier transform with the elements

$$\hat{\varepsilon}^*_{2j-1} = \Big(\frac{2}{n}\Big)^{1/2} \sum_{i=1}^{n} \cos\Big(\frac{2\pi ij}{n}\Big)\hat{\varepsilon}_i,$$
$$\hat{\varepsilon}^*_{2j} = \Big(\frac{2}{n}\Big)^{1/2} \sum_{i=1}^{n} \sin\Big(\frac{2\pi ij}{n}\Big)\hat{\varepsilon}_i, \quad j = 1, \ldots, \Big[\frac{n}{2}\Big].$$

When n is odd, an additional term $\hat{\varepsilon}^*_n = (1/\sqrt{n/2}) \sum_{i=1}^{n} \hat{\varepsilon}_i$ is needed. Let $\hat{\sigma}^2$ be a $\sqrt{n}$-consistent estimate of σ^2 under both the null and the alternative hypotheses. Following Fan (1996), we propose the adaptive Neyman test statistic

$$T^*_{AN} = \max_{1 \le m \le n} \frac{1}{\sqrt{2m\hat{\sigma}^4}} \sum_{i=1}^{m} (\hat{\varepsilon}_i^{*2} - \hat{\sigma}^2). \tag{3.5}$$

We normalize the test statistic as

$$T_{AN} = \sqrt{2\log\log n}\,T^*_{AN} - \{2\log\log n + 0.5\log\log\log n - 0.5\log(4\pi)\}.$$

Theorem 5. *If* C1−C6 *hold, under* H_0 *we have*

$$P(T_{AN} < x) \to \exp(-\exp(-x)) \ \textit{as}\ n \to \infty.$$

Fan and Huang (2001) suggested that the range of maximization over m taken as $[1, n/(\log\log n)^4]$ for convenience. By Theorem 5, the critical region $T_{AN} > -\log\{-\log(1-\alpha)\}$ has an asymptotic significance level of α.

When the single-index model is not a linear model, the residuals need to be ordered according to the monotonic order $\boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}_0$ before using the adaptive Neyman test so that large Fourier coefficients of $\eta(\boldsymbol{X}_i^T\hat{\boldsymbol{\beta}}_0) - \hat{Y}_i, i = 1, \ldots, n$, are concentrated on low frequencies. Such ordering does not affect the size of the proposed adaptive Neyman test.

4. Numerical Studies

4.1. Bandwidth selection

By (2.9), the asymptotic integrated mean squared error of $\hat{\eta}(u)$ is

$$h^4 \int \frac{1}{4}\mu_2^2(\eta''(u))^2\omega(u)du + \frac{\sigma^2 \int (\nu_0/f(u))\omega(u)du}{nh},$$

and the optimal global bandwidth is

$$h_{opt} = C(K)\left[\frac{\sigma^2\int(1/f(u))\omega(u)du}{\int\{\eta''(u)\}^2\omega(u)du}\right]^{1/5} n^{-1/5},$$

where $C(K) = (\nu_0/\mu_2^2)^{1/5}$ and $\omega(u)$ is a weight function. With the form of optimal bandwidth, the plug-in method (Ruppert, Sheather and Wand (1995)) can be applied to estimate the bandwidth for implementatio, but this may not be appropriate for testing and for other statistical inference. We use minimizing cross validation (CV) for choosing the bandwidth,

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{i=1}^{n}\Big(Y_i - \hat{\eta}_{-i,h}(\hat{\boldsymbol{\beta}}_{-i,0}^T\boldsymbol{X}_i;\hat{\boldsymbol{\beta}}_0)\Big)^2, \tag{4.1}$$

where $\hat{\eta}_{-i,h}(\cdot)$ and $\hat{\boldsymbol{\beta}}_{-i,0}$ are the estimates of $\eta(\cdot)$ and $\boldsymbol{\beta}$ with the bandwidth h and without the sample $\boldsymbol{X}_i$, respectively. For constructing SCBs, we note that a much smaller bandwidth is needed for reducing the bias effect. In numerical study, we use $\hat{h}_{opt}$, the bandwidth that minimizes (4.1) for the nonparametric testing problem, but the bandwidth $\hat{h} = \hat{h}_{opt}n^{1/5}n^{-1/3} = \hat{h}_{opt}n^{-2/15}$ for constructing SCBs to nearly satisfy the conditions in the theorems.

4.2. Simulation studies for SCBs

In this section, we report on simulation studies to evaluate the proposed method for constructing SCBs. Throughout, we consider the Epanechnikov kernel $K(t) = 0.75(1-t^2)_+$ for estimating the link function.

Example 1. We consider the SIM,

$$Y_i = (\boldsymbol{\beta}_0^T\boldsymbol{X}_i)^2 + \sigma\varepsilon_i, \quad i = 1, \ldots, n, \tag{4.2}$$

where $\boldsymbol{\beta}_0 = (2,1)^T/\sqrt{5}$, the $\boldsymbol{X}_i$ are bivariate $N\Big((2,2)^T, \boldsymbol{I}_2\Big)$, and the ε_i are standard normal.

Example 2. We generated data $(Y_i, \boldsymbol{X}_i), i = 1, \ldots, n$, from a "sine-bump" model by ignoring the linear part in Carroll et al. (1997),

$$Y_i = \sin\left\{\frac{\pi(\boldsymbol{\beta}_0^T\boldsymbol{X}_i - A)}{B - A}\right\} + \sigma\varepsilon_i, \quad i = 1, \ldots, n, \tag{4.3}$$

where $\boldsymbol{\beta}_0 = (1,1,1)^T/\sqrt{3}, A = \sqrt{3}/2 - 1.645/\sqrt{12}, B = \sqrt{3}/2 + 1.645/\sqrt{12}$, the $\boldsymbol{X}_i$ were 3-dimensional vectors with each component independently Uniform$(0,1)$, and the ε_i were from the standard normal distribution.

Table 1. Coverage probabilities based on 2,000 simulations.

		$1-\alpha=0.90$		$1-\alpha=0.95$	
Example	n	$\sigma=0.1$	$\sigma=0.5$	$\sigma=0.1$	$\sigma=0.5$
Example 1	100	0.8890	0.8820	0.9245	0.9200
	200	0.8990	0.8935	0.9500	0.9470
	300	0.9030	0.8985	0.9575	0.9550
Example 2	100	0.8905	0.8855	0.9135	0.9060
	200	0.8960	0.8910	0.9490	0.9320
	300	0.9090	0.9065	0.9510	0.9405

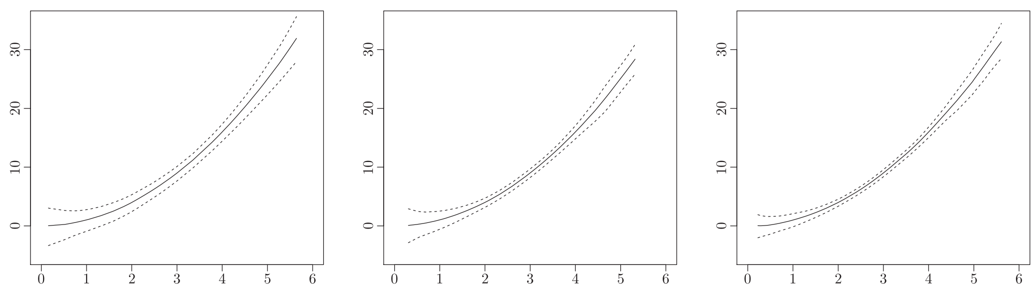

Figure 1. The solid lines are the estimated link function, and the dashed lines are the 95% confidence bands for model (4.2) with $\sigma = 0.1$ and with a sample size at 100, 200, and 300 from left to right.

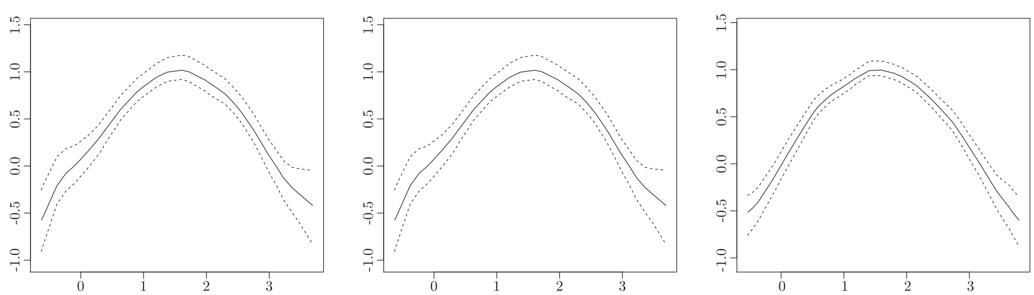

Figure 2. The solid lines are the estimated link function, and the dashed lines are the 95% confidence bands for model (4.3) with $\sigma = 0.1$ and with a sample size at 100, 200, and 300 from left to right.

For the simulations, we took $n = 100$, 200 and 300, and $\sigma = 0.1$ and 0.5, respectively. We conducted 2,000 simulations to compute the average probability of the SCB for the link function in Table 1, and give the 95% pointwise confidence bands of the link function with $\sigma = 0.1$ in Figures 1 and 2. The proposed method works well in these simulated studies.

4.3. Test for SIM

Consider the model

$$Y = X_1 + X_2 + \rho \cdot \exp\{-(X_1 + X_2)^2\} + \sigma\varepsilon, \tag{4.4}$$

where X_1, X_2, and ε are independent standard normals. This is a SIM with link function $\eta(u) = \sqrt{2}u + \rho \cdot \exp(-2u^2)$ and $\beta_{10} = \beta_{20} = \sqrt{2}/2$. We took $n = 100$ and 200, and $\sigma = 0.5$ and 0.8. Based on 1,000 simulations, the scatter plots of Y versus $X_1\hat{\beta}_1 + X_2\hat{\beta}_2$, the true link functions $\eta(\cdot)$ (solid lines) and the estimated link functions $\hat{\eta}(\cdot)$ (dashed lines) with $\sigma = 0.5$ and $n = 200$ are shown in Figure 3 for $\rho = 0, 0.5, 1$ and 1.5. Here $(\hat{\beta}_1, \hat{\beta}_2)^T$ is the profile least squares estimator of $(\beta_1, \beta_2)^T$.

We report the power functions of our proposed test in Figure 4 for all settings at the significance level 0.05. When $\rho = 0$, the simulated power is the type I error. The proposed Neyman test provides a good control on the type I error, and achieves good power. For comparison, Figure 4 gives the power functions of the generalized likelihood ratio test proposed in Fan, Zhang, and Zhang (2001). Our method works better than the generalized likelihood ratio test under most model settings.

5. Application to Data

5.1. Bank data

We consider the bank data in Albright, Winston, and Zappe (1999). The Fifth National Bank of Springfield faced a gender discrimination suit assessment female employees received substantially smaller salaries than male employees. This was a 1995 case with only the bank name changed. These were of 208 employees with complete information on 8 variables. The data set has been analyzed for some high dimensional semiparametric models by Fan and Peng (2004), Lam and Fan (2008), and Li, Lin, and Zhu (2012). To illustrate the proposed method, we consider the SIM,

$$\begin{aligned} \text{Salary} = \eta\Big(&\beta_1\text{Age} + \beta_2\text{YrsExp} + \beta_3\text{Female} + \beta_4\text{PCJob} \\ &+ \sum_{i=1}^{4}\beta_{4+i}\text{Edu}_i + \sum_{i=1}^{5}\beta_{8+i}\text{JobGrd}_i\Big) + \varepsilon, \end{aligned} \tag{5.1}$$

where the variable PCJoB indicates computer related, and the others have evident meaning.

We used the estimation method in Cui, Härdle, and Zhu (2011). The CV method was employed to select the bandwidth, which resulted in $h_{\text{CV}} = 1.435$.

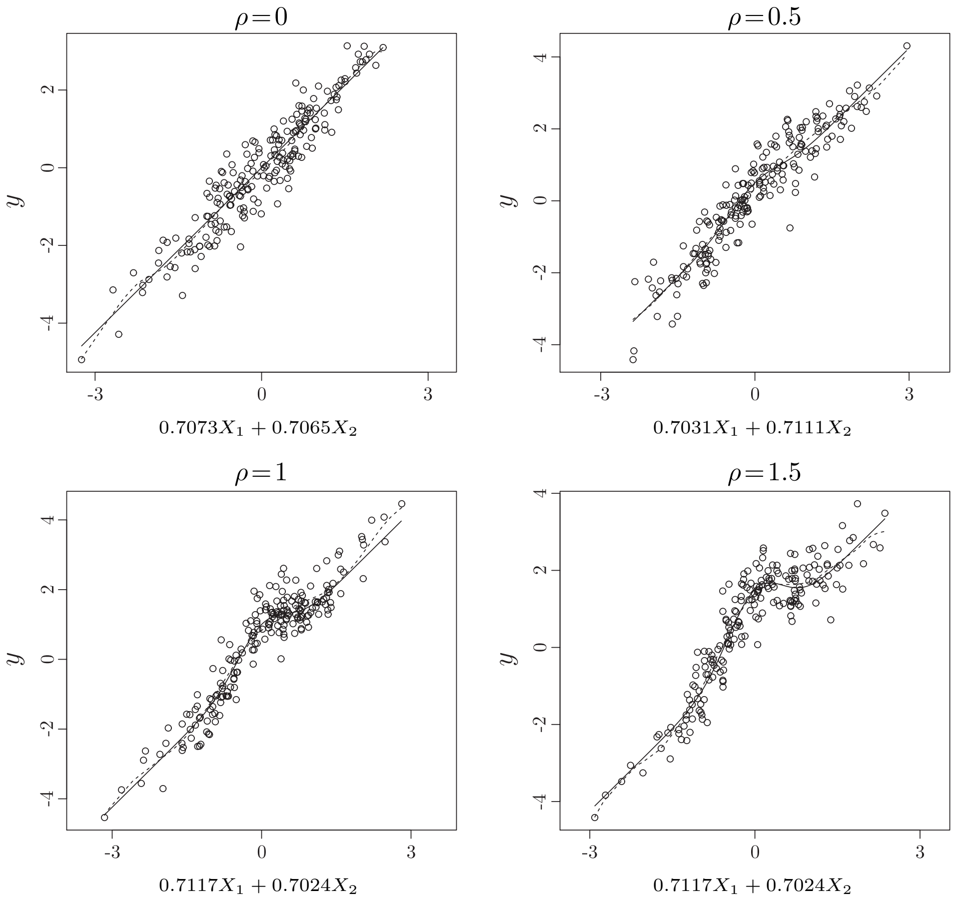

Figure 3. The scatter plots of Y versus $X_1\hat{\beta}_1 + X_2\hat{\beta}_2$. The solid lines denote the true link function $\eta(\cdot)$, and the dashed lines denote the estimated link function $\hat{\eta}(\cdot)$ with $\sigma = 0.5$ and $n = 200$ for $\rho = 0, 0.5, 1$, and 1.5.

Based on the profile least squares, the estimated coefficients and the corresponding standard errors are shown in Table 2. Using the proposed method, we obtained the estimated curve and the 95% confidence band for the link function $\eta(\cdot)$ that are shown in Figure 5(a). Based on the proposed Neyman test, we get the p-value of the proposed test statistic as 0.031;the p-value of the generalized likelihood ratio statistic was 0. This suggests that we reject the null hypothesis that the linear model is the link function. From the residual figure Figure 6(a), there is no special trend for the residuals, suggesting that the SIM is an appropriate model here.

We used all 208 samples to fit (5.1), whereas Fan and Peng (2004) deleted the samples with age over 60 or working experience over 30 and used only 199 samples to fit linear regression model. They did not find strong evidence for

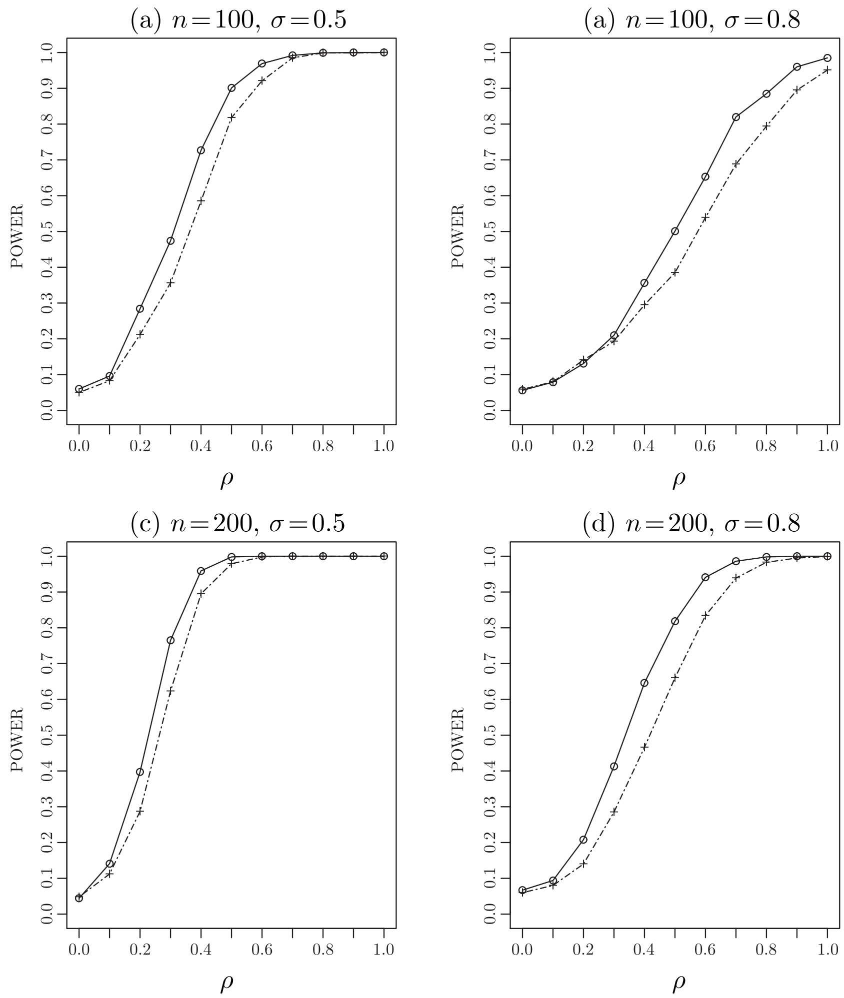

Figure 4. The plots of the estimated power functions for significance level $\alpha = 0.05$. Solid line: Adaptive Neyman test statistic, Dot-dash line: Generalized likelihood ratio test statistic.

discrimination. We believe that the SIM is an appropriate model by using all of the data in view of the residual, Figure 6(a). Figure 5(a) indicates that the link function is a decreasing nonlinear function and, from Table 2, the coefficient of the predictor Female is significant. There is strong evidence of discrimination.

5.2. Car price data

We considered the car price data in Naik and Tsai (2001) and Li, Zhu, and Zhu (2010). There are 25 brands of family sedans in the United States. These brands differ on nine attributes measured by the United States Consumers Union:

Table 2. The estimated coefficients and standard errors for model (5.1).

Variables	Estimator	SE
Age	0.060959239	0.01046973
YrsExp	-0.800956166	0.06471459
Female	0.416505407	0.13657645
PCJob	-0.015036591	0.24988014
Edu_1	0.108489189	0.27227410
Edu_2	0.250099943	0.26133544
Edu_3	-0.009204456	0.15759227
Edu_4	-0.022231789	0.37445128
JobGrd_1	0.221958888	0.37358015
JobGrd_2	0.125374153	0.32750750
JobGrd_3	0.134480046	0.28990072
JobGrd_4	-0.072627564	0.30559367
JobGrd_5	-0.133444393	0.29998355

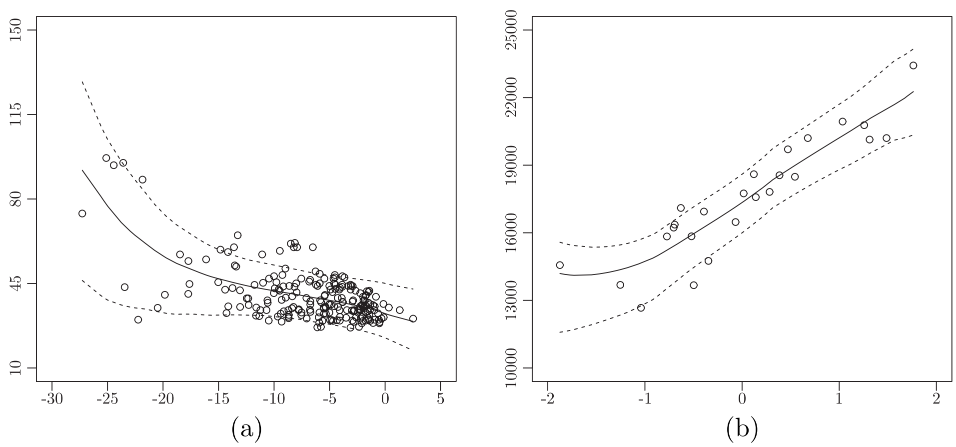

Figure 5. The scatter plot of Y versus $\boldsymbol{X}^T\hat{\boldsymbol{\beta}}$. (a) The solid line is the estimated link function, and the dashed lines are the 95% confidence bands for model (5.1). (b) The solid line is the estimated link function, and the dashed lines are the 95% confidence bands for model (5.2).

mileage per gallon X_1, horsepower X_2, length X_3, width X_4, weight X_5, height X_6, satisfaction X_7, reliability X_8, and overall evaluation X_9. The response variable Y is the non-negotiable transaction price. We applied the Box-Cox transformations on the predictor variables, and consider the following model,

$$Y = \eta\Big(\sum_{i=1}^{9} \beta_i X_i\Big) + \varepsilon. \tag{5.2}$$

The estimation method by Cui, Härdle, and Zhu (2011) was used. The CV method was employed to select the bandwidth, resulting in $h_{\text{CV}} = 1.097$.

Table 3. The estimated coefficients and standard errors for model (5.2).

Parameter	Estimator	SE
β_1	0.034478025	0.0001509354
β_2	0.508355909	0.0001417804
β_3	-0.569947159	0.0001150935
β_4	-0.009204028	0.0002873957
β_5	0.565855995	0.0002455099
β_6	0.068772020	0.0001465410
β_7	0.297566323	0.0002295497
β_8	-0.040345005	0.0001121568
β_9	0.019105593	0.0001655526

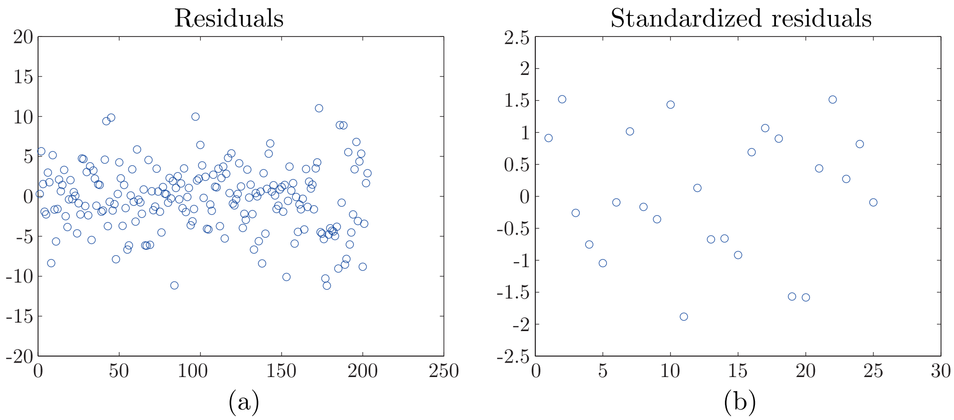

Figure 6. (a) Residuals after fitting the single-index model (5.1). (b) Standardized residuals after fitting the single-index model (5.2).

We report the estimated coefficients and the corresponding standard errors in Table 3. Using our method, the scatter plots of the response Y versus $\hat{\boldsymbol{\beta}}^T\boldsymbol{X}$, the estimated link function and the 95% confidence band are shown in Figure 5(b), and the scatter plot of the standardized residuals is shown in Figure 6(b). From Table 3, we see that horsepower X_2, length X_3, weight X_5 and satisfaction X_7 affect the car price significantly, and length X_3 has a negative effect on the car price. The p-value of the proposed test statistic was 0.660, and the p-value of the generalized likelihood ratio test statistic was 0.221. This, together with the estimated link function in Figure 5(b), suggests that the link function can be approximated by the linear model here.

6. Concluding Remarks

We have derived asymptotic distributions for the link function and its derivative in single-index models and illustrated its usefulness for statistical inference. We have also constructed SCBs for the link function and its derivative in the presence of a nuisance index parameter. We have conducted simulation studies and analyzed two data examples to illustrate the proposed method and the theoretical findings. The methodology in this paper is general and widely applicable. It can be applied to construct SCBs for other functional objects as well. For instance, it can be adapted to the estimation of regression functions in the functional linear model. We expect further research along these lines to yield results with interesting applications.

Acknowledgement

Gaorong Li's research was supported by the NSFC (11101014, 11201306), the Natural Science Foundation of Beijing (1142002), the Science and Technology Project of Beijing Municipal Education Commission (KM201410005010), the Specialized Research Fund for the Doctoral Program of Higher Education of China (20101103120016), PHR(IHLB, PHR20110822) and Program for JingHua Talents in Beijing University of Technology. Heng Peng's research was supported by CERG grants from the Hong Kong Research Grants Council (HKBU 201610, HKBU 201809 and HKBU 202012), FRG grants from Hong Kong Baptist University (FRG2/11-12/130 and FRG2/12-13/077, and a grant from the NSFC (11271094). Tiejun Tong's research was supported by Hong Kong RGC grant HKBU 202711. The authors would like to thank the Editor, an associate editor, and a referee for helpful comments that helped to improve an earlier version of this article.

References

Albright, S. C., Winston, W. L. and Zappe, C. J. (1999). *Data Analysis and Decision Making with Microsoft Excel*. Duxbury, Pacific Grove, CA.

Bickel, P. L. and Rosenblatt, M. (1973). On some global measures of the derivations of density function estimates. *Ann. Statist.* **1**, 1071-1095.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92**, 477-489.

Chen, J., Gao, J. and Li, D. (2013). Estimation in a single-index panel data model with heterogeneous link functions. *Econometric Rev.* **33**, 928-955.

Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31**, 1852-1884.

Cui, X., Härdle, W. K. and Zhu, L. X. (2011). The EFM approach for single-index models. *Ann. Statist.* **39**, 1658-1688.

Eubank, R. L. and Speckman, P. L. (1993), Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88**, 1287-1301.

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

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications.* Chapman & Hall, London.

Fan, J. and Huang, L.-S. (2001). Goodness-of-fit tests for parametric regression models. *J. Amer. Statist. Assoc.* **96**, 640-652.

Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32**, 928-961.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29**, 153-193.

Fan, J. and Zhang, W. Y. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Statist.* **27**, 715-731.

Faraway, J. and Sun, J. (1995). Simultaneous confidence bands for linear regression with heteroscedastic errors. *J. Amer. Statist. Assoc.* **90**, 1094-1098.

Härdle, W. and Bowman, A. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83**, 102-110.

Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21**, 157-178.

Härdle, W. and Stoker, T. M. (1989). Investigating smooth multiple regression by method of average derivatives. *J. Amer. Statist. Assoc.* **84**, 986-995.

Krivobokova, T., Kneib, T. and Claeskens, G. (2010). Simultaneous confidence bands for penalized spline estimators. *J. Amer. Statist. Assoc.* **105**, 852-863.

Lam, C. and Fan, J. (2008). Profile-Kernel likelihood inference with diverging number of parameters. *Ann. Statist.* **36**, 2232-2260.

Li, G. R., Lin, L. and Zhu, L. X. (2012). Empirical likelihood for varying coefficient partially linear model with diverging number of parameters. *J. Multivariate Anal.* **105**, 85-111.

Li, G. R., Zhu, L. X., Xue, L. G. and Feng, S. Y. (2010). Empirical likelihood inference in partially linear single-index models for longitudinal data. *J. Multivariate Anal.* **101**, 718-732.

Li, G. R., Zhu, L. P. and Zhu, L. X. (2010). Adaptive confidence region for the direction in semiparametric regressions. *J. Multivariate Anal.* **101**, 1364-1377.

Liang, H., Liu, X., Li, R. and Tsai, C. L. (2010). Estimation and testing for partially linear single-index models. *Ann. Statist.* **38**, 3811-3836.

Naik, P. and Tsai, C. L. (2001). Single-index model selections. *Biometrika* **88**, 821-832.

Neumann, M. and Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *J. Multivariate Anal.* **9**, 307-333.

Ruppert, D., Sheather, S. J. and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90**, 1257-1270.

Sun, J. and Loader, C. (1994). Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.* **22**, 1328-1345.

Tong, T. J. and Wang, Y. D. (2005). Estimating residual variance in nonparametric regression using least squares. *Biometrika* **92**, 821-830.

Wang, J. L., Xue, L. G., Zhu, L. X. and Chong, Y. (2010). Estimation for a partial-linear single-index model. *Ann. Statist.* **38**, 246-274.

Weisberg, S. and Welsh, A. H. (1994). Adapting for the missing link. *Ann. Statist.* **22**, 1674-1700.

Xia, Y., (1998). Bias-corrected confidence bands in nonparametric regression. *J. Roy. Statist. Soc. Ser. B* **60**, 797-811.

Xia, Y. and Li, W. K. (1999). On single-index coefficient regression models. *J. Amer. Statist. Assoc.* **94**, 1275-1285.

Xia, Y., Li, W. K., Tong, H. and Zhang, D. (2004). A goodness-of-fit test for single-index models. *Statistica Sinica* **14**, 1-39.

Xia, Y., Tong, H., Li, W. K. and Zhu, L. X. (2002). An adaptive estimation of dimension reduction space (with discussion). *J. Roy. Statist. Soc. Ser. B* **64**, 363-410.

Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single index models. *J. Amer. Statist. Assoc.* **97**, 1042-1054.

Zhang, W. Y., Fan, J. Q. and Sun, Y. (2009). A semiparametric model for cluster data. *Ann. Statist.* **37**, 2377-2408.

Zhang, W. Y. and Peng, H. (2010). Simultaneous confidence band and hypothesis test in generalized varying-coefficient models. *J. Multivariate Anal.* **101**, 1656-1680.

Zhu, L. X. and Xue, L. G. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *J. Roy. Statist. Soc. Ser. B* **68**, 549-570.

College of Applied Sciences, Beijing University of Technology, Beijing 100124, P. R. China.

E-mail: ligaorong@gmail.com

Department of Mathematics, Hong Kong Baptist University, Hong Kong.

E-mail: hpeng@math.hkbu.edu.hk

Department of Mathematics, Hong Kong Baptist University, Hong Kong.

E-mail: 12466492@life.hkbu.edu.hk

Department of Mathematics, Hong Kong Baptist University, Hong Kong.

E-mail: tongt@hkbu.edu.hk

(Received April 2012; accepted April 2013)