# Supplementary Material for "A Projection-based Consistent Test Incorporating Dimension-reduction in Partially Linear Models"

Zhihua Sun[1], Feifei Chen[2], Hua Liang[3] and David Ruppert[4]

[1] *University of Chinese Academy of Sciences*

[2] *Beijing Normal University at Zhuhai*

[3] *George Washington University and* [4] *Cornell University*

In this document, we present the detailed proofs of Theorems 1–5, and additional simulation studies.

## S1 Proofs of Theorems 1–5

### S1.1 Proof of Theorem 1

Theorem 1 can be proved in a similar way to prove Theorem 2. We omit the details. □

We list a lemma needed for proving Theorem 2.

**Lemma 1.** *Assuming that Conditions (C1)–(C5) hold, under local alter-*

*native (4.3), we have*

$$\sqrt{n}(\hat{\beta}_n - \beta) = \frac{1}{\sqrt{n}}\Sigma^{-1}\sum_{i=1}^{n}\varepsilon_i\widetilde{X}_i + \Sigma^{-1}\mathrm{E}\{\widetilde{X}\widetilde{D}(X,T)\} + o_p(1).$$

*Proof* . By the definition of $S_j(t,h), j = 0, 1, 2$, the law of large numbers and Conditions (C4)–(C5), we can show that $S_0(t,h) = f_t(t) + o_p(1)$, $S_1(t,h) = h_n^2 f_t'(t)\int z^2K(z)dz + o_p(h_n^2)$, and $S_2(t,h) = h_n^2 f_t(t)\int z^2K(z)dz + o_p(h_n^2)$. Further we can prove that

$$\hat{g}_{1n}(t) = 1/n\sum_{j=1}^{n}X_jK_h(t-T_j)/f_t(t) + o_p(1) = \mathrm{E}(X|T=t) + o_p(1) \quad \text{(S1.1)}$$

and

$$\hat{g}_{2n}(t) = 1/n\sum_{j=1}^{n}Y_jK_h(t-T_j)/f_t(t) + o_p(1) = \mathrm{E}(Y|T=t) + o_p(1). \quad \text{(S1.2)}$$

Thus it is easy to derive that

$$1/n\sum_{i=1}^{n}\{X_i - \hat{g}_{1n}(T_i)\}\{X_i - \hat{g}_{1n}(T_i)\}^\top = \Sigma + o_p(1).$$

This, together with the definition of $\hat{\beta}_n$, yields

$$\sqrt{n}(\hat{\beta}_n - \beta) = \Sigma^{-1}A_n + o_p(1), \quad \text{(S1.3)}$$

where $A_n = 1/\sqrt{n}\sum_{i=1}^{n}\left\{X_i - \hat{g}_{1n}(T_i)\right\}\Big[Y_i - \hat{g}_{2n}(T_i) - \{X_i - \hat{g}_{1n}(T_i)\}^\top\beta\Big]$. For $A_n$, we can decompose it into three parts:

$$A_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{X}_i\Big[\{Y_i - g_2(T_i)\} - \{X_i - g_1(T_i)\}^\top\beta\Big]$$

$$
\begin{aligned}
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{X}_i\Big[g_2(T_i)-\hat{g}_{2n}(T_i)+\{\hat{g}_{1n}(T_i)-g_1(T_i)\}^{\top}\beta\Big]\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{g_1(T_i)-\hat{g}_{1n}(T_i)\Big\}\Big[\{Y_i-\hat{g}_{2n}(T_i)\}-\{X_i-\hat{g}_{1n}(T_i)\}^{\top}\beta\Big]\\
=:\ & A_{n1}+A_{n2}+A_{n3}. \qquad \text{(S1.4)}
\end{aligned}
$$

Note that under $\mathcal{H}_{1n}$ as shown in (4.3), we have $\{Y_i-g_2(T_i)\}-\widetilde{X}_i^{\top}\beta=\varepsilon_i+1/\sqrt{n}\widetilde{D}(X_i,T_i)$, for $i=1,\ldots,n$. It follows from the law of large numbers that

$$
A_{n1}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\widetilde{X}_i+\mathrm{E}\{\widetilde{X}\widetilde{D}(X,T)\}+o_p(1).
$$

For $A_{n2}$, recalling that

$$
\hat{g}_{1n}(t)=\frac{1}{n}\sum_{j=1}^{n}X_jK_h(t-T_j)/f_t(t)+o_p(1),
$$

$$
\hat{g}_{2n}(t)=\frac{1}{n}\sum_{j=1}^{n}Y_jK_h(t-T_j)/f_t(t)+o_p(1),
$$

we have the following decomposition:

$$
\begin{aligned}
A_{n2}\ =\ & n^{-3/2}\sum_{i=1}^{n}\widetilde{X}_i\sum_{j=1}^{n}\Big[g_2(T_j)-Y_j+\{X_j-g_2(T_j)\}^{\top}\beta\Big]K_h(T_i-T_j)/f_t(T_i)\\
&+n^{-3/2}\sum_{i=1}^{n}\widetilde{X}_i\sum_{j=1}^{n}\Big\{g_1(T_j)-g_1(T_i)\Big\}^{\top}\beta K_h(T_i-T_j)/f_t(T_i)\\
&+n^{-3/2}\sum_{i=1}^{n}\widetilde{X}_i\sum_{j=1}^{n}\Big\{g_2(T_i)-g_2(T_j)\Big\}K_h(T_i-T_j)/f_t(T_i)+o_p(1)\\
=:\ & A_{n2}^{[1]}+A_{n2}^{[2]}+A_{n2}^{[3]}+o_p(1).
\end{aligned}
$$

We can prove $A_{n12}^{[j]}=o_p(1),\quad j=1,2,3,$ by proving their second moments converging to zero. So we have

$$A_{n2} = o_p(1). \tag{S1.5}$$

Furthermore, using the similar method in Wang and Sun (2007), we can prove that

$$A_{n3} = o_p(1). \tag{S1.6}$$

From (S1.3)–(S1.6), we can obtain

$$\sqrt{n}(\hat{\beta}_n - \beta) = \frac{\Sigma^{-1}}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_i \widetilde{X}_i + \Sigma^{-1} \mathrm{E}\left\{\widetilde{X}\widetilde{D}(X,T)\right\} + o_p(1). \tag{S1.7}$$

Then Lemma 1 is proved. □

## S1.2 Proof of Theorem 2

For any measurable function $\Upsilon(X, U, W, u)$, let $E\{\Upsilon(X, U, w, u)\} = E\{\Upsilon(X, U, W, u)|W = w\}$ for $w \in \mathbb{S}^{p+1}$. By the definition of $M_{n,pro}(u, W)$, for any given nuisance parameter $W = w \in \mathbb{S}^{p+1}$, we have

$$\begin{aligned}
M_{n,pro}(u,w) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[Y_i - \{X_i^\top \hat{\beta}_n + \hat{g}_n(T_i)\}\right] I(U_i^\top w \le u) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{Y_i - X_i^\top \beta - g(T_i)\right\} I(U_i^\top w \le u) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i^\top (\beta - \hat{\beta}_n) I(U_i^\top w \le u) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{g(T_i) - \hat{g}_n(T_i)\right\} I(U_i^\top w \le u) \\
&=: B_{n1}(u,w) + B_{n2}(u,w) + B_{n3}(u,w).
\end{aligned}$$

For model (4.3), it is easy to prove that

$$B_{n1}(u,w) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i I(U_i^\top w \le u) + E\big\{D(X,T)I(U^\top w \le u)\big\} + o_p(1)$$

and

$$\begin{aligned} B_{n2}(u,w) &= -\frac{1}{\sqrt{n}}\mathrm{E}\{X^\top I(U^\top w \le u)\}\Sigma^{-1}\sum_{i=1}^{n}\varepsilon_i\widetilde{X}_i \\ &\quad -E\{X^\top I(U^\top w \le u)\}\Sigma^{-1}\mathrm{E}\{\widetilde{X}\widetilde{D}(X,T)\} + o_p(1). \end{aligned}$$

For $B_{n3}(u,w)$, we can divide it into

$$\begin{aligned} B_{n3}(u,w) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[g(T_i) - \{\hat{g}_{2n}(T_i) - \hat{g}_{1n}(T_i)^\top\hat{\beta}_n\}\Big] I(U_i^\top w \le u) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[g(T_i) - \{\hat{g}_{2n}(T_i) - \hat{g}_{1n}(T_i)^\top\beta\}\big] I(U_i^\top w \le u) \\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big\{\hat{g}_{1n}(T_i) - g_1(T_i)\big\}^\top(\hat{\beta}_n - \beta)I(U_i^\top w \le u) \\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_1(T_i)^\top(\hat{\beta}_n - \beta)I(U_i^\top w \le u) \\ &=: B_{n3,1}(u,w) + B_{n3,2}(u,w) + B_{n3,3}(u,w). \end{aligned} \tag{S1.8}$$

For $B_{n3,1}(u,w)$, it follows from (S1.1)– (S1.2) that

$$\begin{aligned} B_{n3,1}(u,w) &= n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\big\{g(T_i) - Y_j + X_j^\top\beta\big\}K_h(T_i - T_j)I(U_i^\top w \le u)/f_t(T_i) + o_p(1) \\ &= n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\big\{g(T_j) - Y_j + X_j^\top\beta\big\}K_h(T_i - T_j)I(U_i^\top w \le u)/f_t(T_i) \\ &\quad + n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{g(T_i) - g(T_j)\}K_h(T_i - T_j)I(U_i^\top w \le u)/f_t(T_i) + o_p(1) \\ &=: B_{n3,1}^{[1]}(u,w) + B_{n3,1}^{[2]}(u,w) + o_p(1). \end{aligned}$$

For the first term $B^{[1]}_{n3,1}(u,w)$, we have

$$\begin{aligned} B^{[1]}_{n3,1}(u,w) &= -n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\varepsilon_j K_h(T_i-T_j)I(U_i^\top w\le u)/f_t(T_i) \\ &\quad -n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}D(X_i,T_i)K_h(T_i-T_j)I(U_i^\top w\le u)/f_t(T_i) \\ &=: B^{[1,1]}_{n3,1}(u,w)-B^{[1,2]}_{n3,1}(u,w). \end{aligned}$$

For the first term, it can be proved that

$$\begin{aligned} B^{[1,1]}_{n3,1}(u,w) &= -n^{-3/2}\sum_{j=1}^{n}\varepsilon_j\sum_{i=1}^{n}K_h(T_i-T_j)I(U_i^\top w\le u)/f_t(T_i) \\ &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i \mathrm{E}\{I(U^\top w\le u)|T_i\}+o_p(1). \end{aligned}$$

For the second term, we can validate that

$$\begin{aligned} B^{[1,2]}_{n3,1}(u,w) &= \frac{1}{n}\sum_{i=1}^{n}I(U_i^\top w\le u)/f_t(T_i)\Big\{1/(nh_n)\sum_{j=1}^{n}D(X_i,T_i)K((T_i-T_j)/h_n)\Big\} \\ &= \frac{1}{n}\sum_{i=1}^{n}I(U_i^\top w\le u)\mathrm{E}\{D(X,T)|T_i\}+o_p(1) \\ &= \mathrm{E}\big[I(U^\top w\le u)\mathrm{E}\{D(X,T)|T\}\big]+o_p(1). \end{aligned}$$

Therefore we have

$$B^{[1]}_{n3,1}(u,w) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i \mathrm{E}\left\{I(U^\top w\le u)|T_i\right\} - \mathrm{E}\left[I(U^\top w\le u)\mathrm{E}\{D(X,T)|T\}\right]+o_p(1).$$

For $B^{[2]}_{n3,1}(u,w)$, we have

$$\begin{aligned} B^{[2]}_{n3,1}(u,w) &= n^{-3/2}\sum_{i=1}^{n}I(U_i^\top w\le u)/f_t(T_i)\sum_{j=1}^{n}\{g(T_i)-g(T_j)\}K_h(T_i-T_j) \\ &= n^{-3/2}\sum_{i=1}^{n}I(U_i^\top w\le u)/f_t(T_i)\sum_{j=1}^{n}(T_i-T_j)g'(T_j)K_h(T_i-T_j)+o_p(1) \end{aligned}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} I(U_i^\top w \le u)/f_t(T_i)E\Big\{(T_i - T)g'(T)K_h(T_i - T)|T_i\Big\} + o_p(1).$$

By some tedious proof, we can validate that $B_{n3,1}^{[2]}(u,w) = O_p(\sqrt{n}h_n^2) = o_p(1)$ by Condition (C5). Thus, it yields

$$\begin{aligned} B_{n3,1}(u,w) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i \mathrm{E}\left\{I(U^\top w \le u)|T_i\right\} \\ &\quad -\mathrm{E}\left[I(U^\top w \le u)\mathrm{E}\{D(X,T)|T\}\right] + o_p(1). \quad \text{(S1.9)} \end{aligned}$$

In the following, we consider $B_{n3,2}(u,w)$ and $B_{n3,3}(u,w)$. Following the result of Theorem 6 in Masry (1996), we have that $\sup_t |\hat{g}_{1n}(t) - g_1(t)| = O_P((nh_n)^{-1/2}) + O_P(h_n^2)$. Further note that even under $\mathcal{H}_{1n}$ as shown in (4.3), $\sqrt{n}(\hat{\beta}_n - \beta) = O_P(1)$. Then by the law of large numbers and Condition (C5), it is easy to prove that

$$\begin{aligned} B_{n3,2}(u,w) &= \frac{1}{n}\sum_{i=1}^{n}\{\hat{g}_{1n}(T_i) - g_1(T_i)\}^\top I(U_i^\top w \le u)\sqrt{n}(\hat{\beta}_n - \beta) \\ &= o_p(1), \quad \text{(S1.10)} \end{aligned}$$

and

$$\begin{aligned} B_{n3,3}(u,w) &= \mathrm{E}\{g_1(T)^\top I(U^\top w \le u)\}\sqrt{n}(\hat{\beta}_n - \beta) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\mathrm{E}\{g_1(T)^\top I(U^\top w \le u)\}\Sigma^{-1}\sum_{i=1}^{n}\varepsilon_i \widetilde{X}_i \\ &\quad +\mathrm{E}\{g_1(T)^\top I(U^\top w \le u)\}\Sigma^{-1}\mathrm{E}\{\widetilde{X}\widetilde{D}(X,T)\} \\ &\quad +o_p(1). \quad \text{(S1.11)} \end{aligned}$$

Therefore by (S1.8)– (S1.11), we have

$$\begin{aligned}
B_{n3}(u,w) &= \frac{1}{\sqrt{n}}\mathrm{E}\{g_1(T)^\top I(U^\top w\le u)\}\Sigma^{-1}\sum_{i=1}^n \varepsilon_i \widetilde{X}_i \\
&\quad -\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i \mathrm{E}\{I(U^\top w\le u)|T_i\} \\
&\quad +\mathrm{E}\{g_1(T)^\top I(U^\top w\le u)\}\Sigma^{-1}\mathrm{E}\{\widetilde{X}\widetilde{D}(X,T)\} \\
&\quad -\mathrm{E}\left\{I(U^\top w\le u)\mathrm{E}\{D(X,T)|T\}\right\}+o_p(1).
\end{aligned}$$

So we have the following expression for $M_{n,pro}(u,w)$,

$$\begin{aligned}
M_{n,pro}(u,w) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i[I(U_i^\top w\le u)-\mathrm{E}\{I(U^\top w\le u)|T_i\}] \\
&\quad -\frac{1}{\sqrt{n}}\Gamma(u,w)\Sigma^{-1}\sum_{i=1}^n \varepsilon_i \widetilde{X}_i-\Gamma(u,w)\Sigma^{-1}\mathrm{E}\{\widetilde{X}\widetilde{D}(X,T)\} \\
&\quad +\mathrm{E}\{\widetilde{D}(X,T)I(U^\top w\le u)\}+o_p(1).
\end{aligned}$$

Let $M_u^{[1]}(\tilde{u},w)=I(\tilde{u}^\top w\le u)$. It is easy to see that $M_u^{[1]}(\tilde{u},w)$ is monotone with respect to $u$. By Lemma 9.10 of Kosorok (2008), the function class $\{M_u^{[1]}(\tilde{u},w):u\in\mathbb{R}^1\}$ is a VC-class. Let $M_u^{[2]}(\tilde{u},w)=\mathrm{E}\{I(U^\top w\le u)|T,W=w\}$. Then similarly the function classes $\{M_u^{[2]}(\tilde{u},w):u\in\mathbb{R}^1\}$ and $\{\Gamma(u,w):u\in\mathbb{R}^1\}$ are both VC-classes. By Lemma 2.6.8 of van der Vaart and Wellner (1996), the function classes $\{\varepsilon M_u^{[1]}(\tilde{u},w):u\in\mathbb{R}^1\}$, $\{\varepsilon M_u^{[2]}(\tilde{u},w):u\in\mathbb{R}^1\}$, and the class $\{\Gamma(u,w)\Sigma^{-1}\varepsilon\widetilde{X}:u\in\mathbb{R}^1\}$ are all VC-class. Then by Lemma 9.17 of Kosorok (2008), the function class $\{\Psi_u(\tilde{u},\tilde{y},\varepsilon,w):u\in\mathbb{R}^1\}$ is a VC-class. We can take the envelop function as $|\varepsilon|+\mathrm{E}(\|\widetilde{X}^\top\|)|\varepsilon|\Sigma^{-1}|\widetilde{X}|$. By Theorem 2.6.7 and Theorem 2.5.2 of van der

Vaart and Wellner (1996), we can prove that for given $W = w$, the estimated empirical process $M_{n,pro}(u, w)$ converges weakly to $M_{pro}(u, w) + \Omega(u, w)$ in the Skorohod space $S[-\infty, \infty]$. It can further be obtained the result that $M_{n,pro}(u, w)$ converges weakly to $M_{pro}(u, w) + \Omega(u, w)$ in the Skorohod space $S[\Pi]$ by the fact that $W$ is independent of $(X, Y, T, \varepsilon)$ and follows the uniform distribution on the unit ball in $\mathbb{R}^{p+1}$. By the continuous mapping theorem, we can prove the result for $\mathcal{T}_{n,pro}$. $\square$

## S1.3 Proof of Theorem 3

Under the local alternative (4.4), we have

$$\sqrt{n}(\hat{\beta}_n - \beta) = \frac{1}{\sqrt{n}}\Sigma^{-1}\sum_{i=1}^{n}\varepsilon_i\widetilde{X}_i + \sqrt{n}\Sigma^{-1}\mathrm{E}\{\widetilde{X}\widetilde{D}(X, T)\} + o_p(1).$$

By a similar method to prove the results of Theorem 2, we can validate that for any given $W = w$,

$$M_{n,pro}(u, w) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_u(U_i, Y_i, \varepsilon_i, w) + \sqrt{n}\Omega(u, w) + o_p(1).$$

As $n \to \infty$, $\sqrt{n}\Omega(u, w) \to \infty$ for any $u \in \mathbb{R}^1$. Thus $M_{n,pro}(u, w) \to \infty$ as $n \to \infty$. The result of Theorem 3 follows. $\square$

## S1.4 Proof of Theorem 4

The proof of Theorem 4 is similar to that of Theorem 3. We omit the details. $\square$

## S1.5 Proof of Theorem 5

We only prove the result under alternative hypothesis (4.4). By assuming the deviation function $D(X,T)=0$, the result under the null hypothesis (1.1) can be proved similarly and the details are omitted.

Let $\hat\beta_n^* = [\sum_{i=1}^n\{X_i-\hat g_{1n}(T_i)\}\{X_i-\hat g_{1n}(T_i)\}^\top]^{-1}\sum_{i=1}^n\{X_i-\hat g_{1n}(T_i)\}\{Y_i^*-\hat g_{2n}^*(T_i)\}$, where $Y_i^* = X_i^\top\hat\beta_n+\hat g_n(T_i)+[Y_i-\{X_i^\top\hat\beta_n+\hat g_n(T_i)\}]V_i$ and

$$\hat g_{2n}^*(t) = \frac{1}{n}\sum_{i=1}^n\frac{\{S_2(t,h)-S_1(t,h)(T_i-t)\}K_h(t-T_i)Y_i^*}{S_0(t,h)S_2(t,h)-S_1^2(t,h)}.$$

We first prove that

$$\sqrt{n}(\hat\beta_n^*-\hat\beta_n) = \frac{\Sigma^{-1}}{\sqrt{n}}\sum_{i=1}^n\widetilde X_i^* V_i[Y_i-\{X_i^\top\hat\beta_n+\hat g_n(T_i)\}]+o_p(1) \qquad \text{(S1.12)}$$

with $\widetilde X_i^* = X_i-\hat g_{1n}(T_i), i=1,\ldots,n$. It is easy to see that $1/n\sum_{i=1}^n\{X_i-\hat g_{1n}(T_i)\}\{X_i-\hat g_{1n}(T_i)\}^\top = \Sigma+o_p(1)$. Then we have

$$\sqrt{n}(\hat\beta_n^*-\hat\beta_n) = \Sigma^{-1}A_n^*+o_p(1), \qquad \text{(S1.13)}$$

where $A_n^* = 1/\sqrt{n}\sum_{i=1}^n\{X_i-\hat g_{1n}(T_i)\}[Y_i^*-\hat g_{2n}^*(T_i)-\{X_i-\hat g_{1n}(T_i)\}^\top\hat\beta_n]$. For $\hat g_{2n}^*(t)$, we have

$$\begin{aligned}\hat g_{2n}^*(t) \quad &= \quad \frac{1}{n}\sum_{j=1}^n\frac{\{S_2(t,h)-S_1(t,h)(T_j-t)\}K_h(t-T_j)}{S_0(t,h)S_2(t,h)-S_1^2(t,h)}\\ &\qquad \left(X_j^\top\hat\beta_n+\hat g_n(T_j)+[Y_j-\{X_j^\top\hat\beta_n+\hat g_n(T_j)\}]V_j\right).\end{aligned}$$

Note that the i.i.d. random variable sequence $\{V_j, j=1,\ldots,n\}$ with mean zero and variance 1. Further, $V_j$ is independent of $(Y_j,X_j,T_j)$. Then it can be validated that

$$\hat{g}^*_{2n}(t) = \hat{g}_{1n}(t)^\top \hat{\beta}_n + \hat{g}_n(t) + o_p(h_n^2). \tag{S1.14}$$

Thus $Y_i^* - \hat{g}^*_{2n}(T_i) = X_i^\top \hat{\beta}_n + \hat{g}_n(T_i) + \big[Y_i - \{X_i^\top \hat{\beta}_n + \hat{g}_n(T_i)\}\big] V_i - \{\hat{g}_{1n}(T_i)^\top \hat{\beta}_n + \hat{g}_n(T_i)\} + o_p(h_n^2) = \{X_i - \hat{g}_{1n}(T_i)\}^\top \hat{\beta}_n + [Y_i - \{X_i^\top \hat{\beta}_n + \hat{g}_n(T_i)\}] V_i + o_p(h_n^2)$. Then $Y_i^* - \hat{g}^*_{2n}(T_i) - \{X_i - \hat{g}_{1n}(T_i)\}^\top \hat{\beta}_n = [Y_i - \{X_i^\top \hat{\beta}_n + \hat{g}_n(T_i)\}] V_i + o_p(h_n^2)$. For $A_n^*$, we have

$$A_n^* = 1/\sqrt{n} \sum_{i=1}^n \widetilde{X}_i^* V_i [Y_i - \{X_i^\top \hat{\beta}_n + \hat{g}_n(T_i)\}] + o_p(1). \tag{S1.15}$$

Then from (S1.13) and (S1.15), we can prove (S1.12).

Let $M^*_{n,pro}(u,w) = 1/\sqrt{n}\sum_{i=1}^n [Y_i^* - \{X_i^\top \hat{\beta}_n^* + \hat{g}_n^*(T_i)\}] I(U_i^\top w \le u), u \in \mathbb{R}^1$, where $\hat{g}_n^*(t) = \hat{g}^*_{2n}(t) - \hat{g}_{1n}(t)^\top \hat{\beta}_n^*$. In the following, we aim for proving that

$$\begin{aligned} M^*_{n,pro}(u,w) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i [Y_i - \{{X_i}^\top \hat{\beta}_n + \hat{g}_n(T_i)\}][I(U_i^\top w \le u) \\ &\quad - \mathrm{E}\{I(U^\top w \le u)|T_i\}] - \frac{1}{\sqrt{n}} \Gamma(u,w) \Sigma^{-1} \sum_{i=1}^n V_i \\ &\quad \times [Y_i - \{X_i^\top \hat{\beta}_n + \hat{g}_n(T_i)\}] \widetilde{X}_i^* + o_p(1). \end{aligned}$$

Note that $Y_i^* - \{X_i^\top \hat{\beta}_n^* + \hat{g}_n^*(T_i)\} = X_i^\top \hat{\beta}_n + \hat{g}_n(T_i) + [Y_i - \{X_i^\top \hat{\beta}_n + \hat{g}_n(T_i)\}] V_i - \{X_i^\top \hat{\beta}_n^* + \hat{g}_n^*(T_i)\} = [Y_i - \{X_i^\top \hat{\beta}_n + \hat{g}_n(T_i)\}] V_i - X_i^\top (\hat{\beta}_n^* - \hat{\beta}_n) - \{\hat{g}_n^*(T_i) - \hat{g}_n(T_i)\}]$. Then we have

$$M^*_{n,pro}(u,w) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big( [Y_i - \{X_i^\top \hat{\beta}_n + \hat{g}_n(T_i)\}] V_i - X_i^\top (\hat{\beta}_n^* - \hat{\beta}_n)$$

$$
\begin{aligned}
& -\{\hat{g}_n^*(T_i) - \hat{g}_n(T_i)\}\Big) I(U_i^\top w \le u) \\
= \quad & \frac{1}{\sqrt{n}} \sum_{i=1}^n [Y_i - \{X_i^\top \hat{\beta}_n + \hat{g}_n(T_i)\}] V_i I(U_i^\top w \le u) \\
& - \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^\top (\hat{\beta}_n^* - \hat{\beta}_n) I(U_i^\top w \le u) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\hat{g}_n(T_i) - \hat{g}_n^*(T_i)\} I(U_i^\top w \le u) \\
=: \quad & M_{n1,pro}^*(u,w) - M_{n2,pro}^*(u,w) + M_{n3,pro}^*(u,w). \qquad \text{(S1.16)}
\end{aligned}
$$

It follows from (S1.13) that

$$
\begin{aligned}
M_{n2,pro}^*(u,w) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^\top (\hat{\beta}_n^* - \hat{\beta}_n) I(U_i^\top w \le u) \\
&= \frac{1}{\sqrt{n}} \mathrm{E}\{X^\top I(U^\top w \le u)\} \Sigma^{-1} \sum_{i=1}^n \widetilde{X}_i^* V_i \\
&\quad \times [Y_i - \{X_i^\top \hat{\beta}_n + \hat{g}_n(T_i)\}] + o_p(1). \qquad \text{(S1.17)}
\end{aligned}
$$

In the same way, we have

$$
\begin{aligned}
M_{n3,pro}^*(u,w) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n [\hat{g}_n(T_i) - \{\hat{g}_{2n}^*(T_i) - \hat{g}_{1n}(T_i)^\top \hat{\beta}_n^*\}] I(U_i^\top w \le u) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n [\hat{g}_n(T_i) - \{\hat{g}_{2n}^*(T_i) - \hat{g}_{1n}(T_i)^\top \hat{\beta}_n\}] I(U_i^\top w \le u) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{g}_{1n}(T_i)^\top (\hat{\beta}_n^* - \hat{\beta}_n)\} I(U_i^\top w \le u) \\
&=: M_{n3,1,pro}^*(u,w) + \frac{1}{\sqrt{n}} \Gamma(u,w) \Sigma^{-1} \sum_{i=1}^n V_i \\
&\quad \times [Y_i - \{X_i^\top \hat{\beta}_n + \hat{g}_n(T_i)\}] \hat{g}_n(T_i) + o_p(1).
\end{aligned}
$$

From (S1.14), we have

$$M^*_{n3,1,pro}(u,w) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\hat{g}_n(T_i) + \hat{g}_{1n}(T_i)^\top\hat{\beta}_n \\ -\frac{1}{n}\sum_{j=1}^{n}\frac{\{S_2(t,h) - S_1(t,h)(T_j - t)\}K_h(t - T_j)Y_j^*}{S_0(t,h)S_2(t,h) - S_1^2(t,h)}\Big] \times I(U_i^\top w \le u).$$

By the definition of $Y_j^*$, it can be proved that

$$\begin{aligned} M^*_{n3,1,pro}(u,w) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_h(T_i - T_j)}{nf_t(T_i)}[Y_j - \{X_j^\top\hat{\beta}_n + \hat{g}_n(T_j)\}]V_j I(U_i^\top w \le u) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i[Y_i - \{X_i{}^\top\hat{\beta}_n + \hat{g}_n(T_i)\}]\mathrm{E}\{I(U^\top w \le u)|T_i\} + o_p(1). \end{aligned}$$

Therefore, it follows that

$$\begin{aligned} M^*_{n3pro}(u,w) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i[Y_i - \{X_i{}^\top\hat{\beta}_n + \hat{g}_n(T_i)\}]\mathrm{E}\{I(U^\top w \le u)|T_i\} - \frac{1}{\sqrt{n}}\Gamma(u,w)\times \\ &\quad \Sigma^{-1}\sum_{i=1}^{n}V_i[Y_i - \{X_i^\top\hat{\beta}_n + \hat{g}_n(T_i)\}]\hat{g}_n(T_i) + o_p(1). \end{aligned} \tag{S1.18}$$

From (S1.16)–(S1.18), and by the facts that the i.i.d. random variable sequence $\{V_j, j = 1,\ldots,n\}$ with mean zero and variance 1 and that $V_j$ is independent from $(Y_j, X_j, T_j)$, we can further prove that

$$\begin{aligned} M^*_{n,pro}(u,w) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i[Y_i - \{X_i{}^\top\beta + g(T_i)\}]\{I(U_i^\top w \le u) \\ &\quad -\mathrm{E}\{I(U^\top w \le u)|T_i\} - \frac{1}{\sqrt{n}}\Gamma(u,w)\Sigma^{-1}\sum_{i=1}^{n}V_i \\ &\quad \times[Y_i - \{X_i^\top\beta + g(T_i)\}]\widetilde{X}_i + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\Psi_u(U_i, Y_i, \varepsilon_i, w) + o_p(1). \end{aligned}$$

In the above process of proof, we should note that the terms such as $\mathrm{E}\{V\widetilde{D}(X,T)I(U^\top w \le u)\}$ are all zero due to the fact that mean-zero variable $V_j$ is independent of $(Y_j, X_j, T_j)$ for $j = 1, \dots, n$. Moreover, it is easy to prove that for given $w$, the function class $\{V\Psi_u(\tilde{u}, \tilde{y}, \varepsilon, w) : u \in \mathbb{R}^1\}$ is a VC-class. By the similar argument to the proof of Theorem 2, we can prove that the conditional distribution of $\mathcal{T}^*_{n,pro}$ converges in distribution to the limiting null distribution of $\mathcal{T}_{n,pro}$. □

## S2 Additional simulation studies

In this section, we report additional simulation results to evaluate the finite sample performance of the proposed method. The settings are the similar to those in the main text of the article but with 5− and 10− dimensional linear covariates. Two examples are considered:

**Example 1.** The candidate models have 5–dimensional linear covariates and possible interaction between the linear and nonparametric components:

$$Y = X^\top\beta + g(T) + C\sum_r X_r T/2 + \varepsilon, \tag{S2.1}$$

where $X = (X_1, \dots, X_5)^\top$, $g(T) = \sin(T^2) + 2T$, $T \sim \mathcal{N}(0,1)$, $\varepsilon \sim \mathcal{N}(0,1)$, and $\beta = (-2,-1,2,1,3)^\top$. Let $X$ follow multivariate normal distribution $\mathcal{N}_5(0,\Sigma)$ with $\Sigma = (\sigma_{jj'})$ and $\sigma_{jj} = 1, j = 1, \dots, 5; \sigma_{jj'} = \sigma_{j'j} = 0.3, j' =$

$j+1, j=1,2,3,4$. $C$ was chosen to be $0, 0.2, 0.4, 0.6, 0.8, 1.0$.

**Example 2.** The candidate models have 10–dimensional linear covariates and possible high effects between the covariates in linear parts:

$$Y = X^\top \beta + g(T) + C \sum_r X_r^2/10 + \varepsilon, \tag{S2.2}$$

where $X = (X_1, \ldots, X_{10})^\top$, $g(T) = T^3$, $T \sim \mathcal{N}(1,1)$, $\varepsilon \sim \mathcal{N}(0, 0.5)$, and $\beta = (0.3, 0.8, 0.4, -1.3, -1, 0.2, 0.5, -0.2, -0.4, 1.2)^\top$. The linear covariates $X$ follow a multivariate normal distribution $\mathcal{N}_{10}(0, \Sigma)$ with $\Sigma = (\sigma_{jj'})$ and $\sigma_{jj'} = 0.1^{|j-j'|}, j, j' = 1, \ldots, 10$. To consider different data generating processes, we chose $C = 0, 0.5, 1, 2, 3, 4$.

In Example 1, the empirical sizes of $T_n^s$ and $\mathcal{T}_{n,Pro}$ are close to the nominal levels, while the empirical sizes of $T_n^u$ and $T_n^{Xia}$ are lower than the nominal levels. Observed the power curves, $\mathcal{T}_{n,Pro}$ stands out remarkably as the best method, followed by $T_n^s$, $T_n^{Xia}$ and $T_n^u$ in a consistent order for all different configurations.

In Example 2, the performance of the four tests still show a ranking similar to the one in Example 1, except that $T_n^{Xia}$ is slightly superior to $T_n^s$ in some configurations. However, the performance of $\mathcal{T}_{n,Pro}$ is much better than that of the other three tests. $T_n^u$ always has very small empirical size and power no matter how large or small the sample size. $T_n^s$ also yields low

Table 1: Failure times among the 1000 replicates for the four tests in models (S2.1) and (S2.2) with different sample sizes and different $C$ values.

| | n=60 | | n=100 | | n=200 | | | |
|---|---|---|---|---|---|---|---|---|
| C | $F^u$ | $F^s$ | $F^u$ | $F^s$ | $F^u$ | $F^s$ | $F^{Xia}$ | $F^{Pro}$ |
| | | | | model (S2.1) | | | | |
| 0.0 | 156 | 0 | 136 | 0 | 78 | 0 | 0 | 0 |
| 0.2 | 151 | 0 | 116 | 0 | 79 | 0 | 0 | 0 |
| 0.4 | 142 | 0 | 109 | 0 | 70 | 0 | 0 | 0 |
| 0.6 | 154 | 0 | 111 | 0 | 80 | 0 | 0 | 0 |
| 0.8 | 151 | 0 | 99 | 0 | 76 | 0 | 0 | 0 |
| 1.0 | 166 | 0 | 105 | 0 | 73 | 0 | 0 | 0 |
| | | | | model (S2.2) | | | | |
| 0.0 | 998 | 173 | 999 | 12 | 1000 | 0 | 0 | 0 |
| 0.1 | 997 | 170 | 1000 | 14 | 1000 | 0 | 0 | 0 |
| 0.2 | 994 | 163 | 1000 | 9 | 1000 | 0 | 0 | 0 |
| 0.3 | 991 | 160 | 998 | 15 | 998 | 0 | 0 | 0 |
| 0.4 | 998 | 171 | 997 | 8 | 999 | 0 | 0 | 0 |
| 0.5 | 995 | 164 | 999 | 18 | 1000 | 0 | 0 | 0 |

$F^u$, $F^s$, $F^{Xia}$ and $F^{Pro}$: corresponding to the tests by Fan and Li (1996), Zhu and Ng (2003), Xia (2009) and this paper.

empirical size and power, whereas $T_n^{Xia}$ cannot control the type I error.

We report failure times of the tests $T_n^u$, $T_n^s$, $T_n^{Xia}$ and $\mathcal{T}_{n,Pro}$ under Examples 1 and 2 in Table 1. In Example 1, $T_n^s$ did not degenerate, but $T_n^u$ degenerated at least 150 times among the 1000 replicates for sample size $n = 60$, and at least 70 times for sample size $n = 200$. This situation becomes much worse when the dimension of $X$ increases to 10. Both

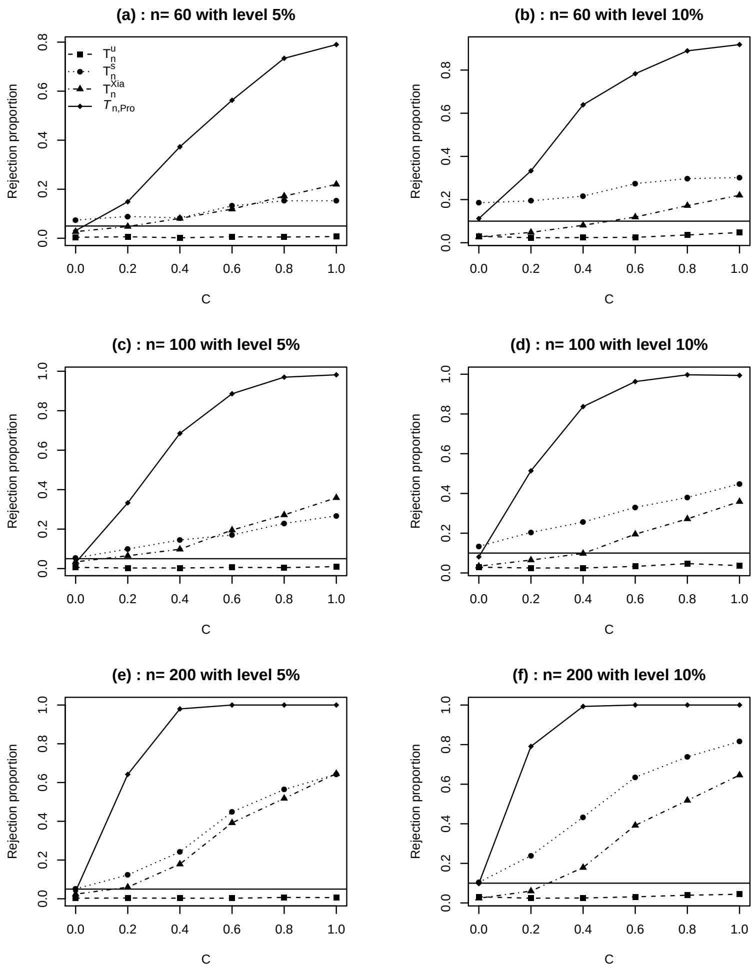


Figure 1: Simulation results for model (S2.1) in Example 1. Rejection proportions of four methods against $C$ with different sample sizes and test levels 0.05, 0.1. $\mathcal{T}_{n,Pro}$: the proposed test (solid line with filled diamond); $T_n^s$: the test of Zhu and Ng (2003) (dotted line with filled circle); $T_n^u$: the test of Fan and Li (1996) (dashed line with filled square); and $T_n^{Xia}$: the test of Xia (2009) (dot-dash line with filled triangle). The thin horizontal line indicates the nominal level 0.05 or 0.1.

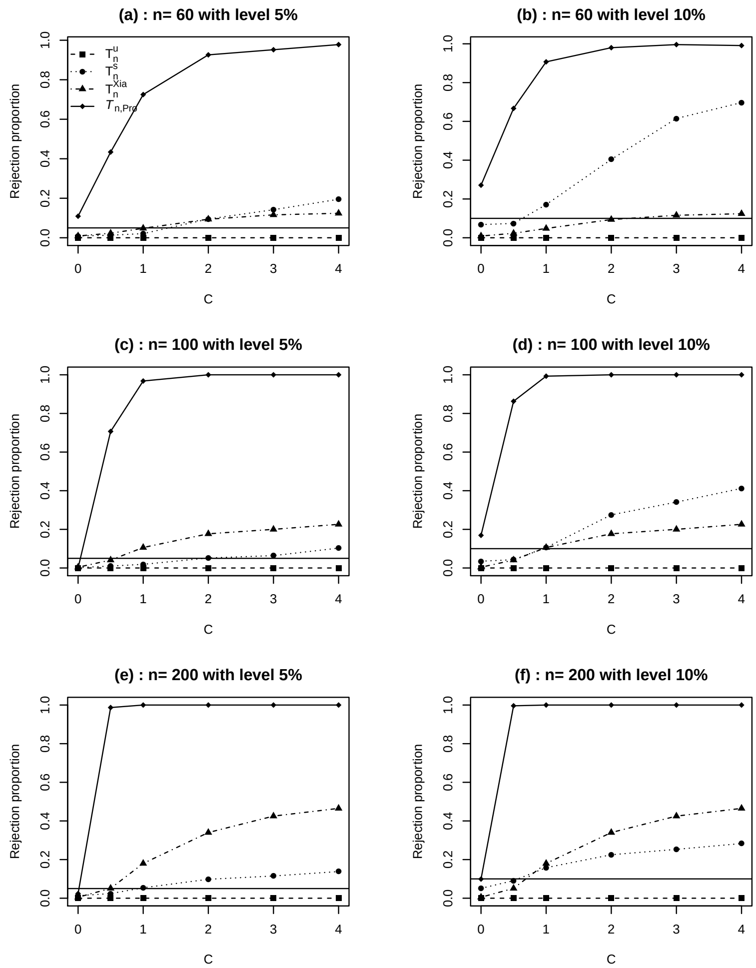


Figure 2: Simulation results for model (S2.2) in example 2. The legend is the same as in Figure 1.

$T_n^u$ and $T_n^s$ almost always degenerated in all configurations. This may explain why the power curves of these two tests in Figures 1 and 2 are so flat. This deficiency for $T_n^u$ is not surprising given the fact that $T_n^u$ is based on the local smoothing estimation for $\mathrm{E}(\varepsilon|X,T)$. For $T_n^s$, recall that

$T_n^s = \int \left\{1/\sqrt{n}\sum_{i=1}^n \hat{\varepsilon}(U_i, \hat{\beta}_n, \hat{g}_n(T_i))\ I(X_i \leq x)I(T_i \leq t)\right\}^2 F_n d(x,t)$. The indicator weighting function $I(X_i \leq x)I(T_i \leq t)$ can easily be null, and cause $T_n^s$ to degenerate. Such a degeneration implicitly indicates that $T_n^s$ still suffers from curse of dimensionality.

## References

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

Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer-Verlag.

Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal Time Series Analysis 17*, 571–599.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. New York: Springer-Verlag.

Wang, Q. and Z. Sun (2007). Estimation in partially linear models with missing responses at random. *Journal of Multivariate Analysis 98*(7), 1470–1493.

Xia, Y. (2009). Model checking in regression via dimension reduction. *Biometrika 96*, 133–148.

Zhu, L. X. and K. W. Ng (2003). Checking the adequacy of a partial linear model. *Statistica Sinica 13*, 763–781.