An Adaptive-to-Model Test for Parametric Single-Index Errors-in-Variables Models

Hira L. Koul, Chuanlong Xie, Lixing Zhu*

Michigan State University, USA

Hong Kong Baptist University, Hong Kong, China

Jinan University, Guangzhou, China

Supplementary Material

The supplement is organized as follows. Section S1 provides more simulations results to compare T_n and $\tilde{T}_n$. In Section S2, Proposition 2 is proved. The proof of Theorem 1 appears in Section S3. Based on the asymptotic behavior of $\hat{\theta}_0$ and $\hat{B}(\hat{q})$ under the local alternatives, the proof of Theorem 5 is included in Section S4. As Theorem 2 is a special case of Theorem 5 when $C_n = 0$, its proof is omitted. In Section S5, we only sketch the proof of Theorem 1 as it is similar to that of Theorem 5. Section S6 shows a sketch of the proof for Theorem 3.

S1 Simulation

The comparison between T_n and $\tilde{T}_n$ is another purpose of **Study 1.** The results are reported in Tables 1.

Tables 1 about here

We find that the empirical power of $\tilde{T}_n$ is slightly higher than that of T_n, but the size of $\tilde{T}_n$ also tends to be slightly larger, even when $n = 200$ and $p = 2$. Although $\tilde{T}_n$ has bias, but each residual in $\tilde{T}_n$ is estimated by all validation data which is more precise with smaller variance than that of T_n derived by half validation dat. We can then conclude, based on this limited simulation, the test $\tilde{T}_n$ is slightly more liberal than the bias-corrected test T_n, but also slightly more powerful. These two tests are competitive.

S2 Proof of Proposition 2

The claim (1) has been proved in Lee and Sepanski (1995). We now prove the claim (2). The estimator is $\hat{\theta}_0 = \arg\min_\theta Q_n(\theta)$ where

$$\begin{aligned} Q_n(\theta) &= \frac{1}{n}\Big(\boldsymbol{Y} - \boldsymbol{D}(\tilde{\boldsymbol{D}}^{\mathrm{T}}\tilde{\boldsymbol{D}})^{-1}\tilde{\boldsymbol{D}}^{\mathrm{T}} g(\tilde{\boldsymbol{X}}\beta, \gamma)\Big)^{\mathrm{T}} \\ &\quad \times\Big(\boldsymbol{Y} - \boldsymbol{D}(\tilde{\boldsymbol{D}}^{\mathrm{T}}\tilde{\boldsymbol{D}})^{-1}\tilde{\boldsymbol{D}}^{\mathrm{T}} g(\tilde{\boldsymbol{X}}\beta, \gamma)\Big). \end{aligned}$$

Here $\tilde{\boldsymbol{X}}$ is the $N \times p$ matrix whose s-th row is $\tilde{x}_s^T$, $s = 1, \cdots, N$, $\boldsymbol{Y}$ is a $n \times 1$ vector, and $g(\tilde{\boldsymbol{X}}\beta, \gamma)$ represents $N \times 1$ vector $[g(\beta^{\mathrm{T}}\tilde{x}_1, \gamma), \cdots, g(\beta^{\mathrm{T}}\tilde{x}_N, \gamma)]^{\mathrm{T}}$. The matrices $\boldsymbol{D}$ and $\tilde{\boldsymbol{D}}$ are design matrices according to g. More precisely, $\boldsymbol{D}$ is the $n \times k$ matrix whose i-th row denoted by $\bar{w}_i^{\mathrm{T}}$, is a vector consisting of polynomials of w_i, while $\tilde{\boldsymbol{D}}$ is the corresponding matrix of validation data, whose s-th row $\bar{\tilde{w}}_s^{\mathrm{T}}$ is a vector consisting of polynomials of $\tilde{w}_s$. For linear model, $\bar{w}_i = w_i$ and $\bar{\tilde{w}}_s = \tilde{w}_s$. For nonlinear model, we let $\bar{w}_i(\bar{\tilde{w}}_s)$ be the vector consisting of a constant and the first two order polynomials of $w_i(\tilde{w}_s)$. Let $\bar{W}$ is the vector consist of polynomials of W. It is easy to see that $Q_n(\theta)$ uniformly converges in probability to

$$\begin{aligned} Q(\theta) &= \left\{E^{-1}[\bar{W}\bar{W}^{\mathrm{T}}]E[\bar{W}Y] - E^{-1}[\bar{W}\bar{W}^{\mathrm{T}}]E[\bar{W}g(\beta^{\mathrm{T}}X,\gamma)]\right\}^{\mathrm{T}} E[\bar{W}\bar{W}^{\mathrm{T}}] \\ &\quad \times \left\{E^{-1}[\bar{W}\bar{W}^{\mathrm{T}}]E[\bar{W}Y] - E^{-1}[\bar{W}\bar{W}^{\mathrm{T}}]E[\bar{W}g(\beta^{\mathrm{T}}X,\gamma)]\right\} \\ &\quad + \{E[Y^2] - E[Y\bar{W}^{\mathrm{T}}]E^{-1}[\bar{W}\bar{W}^{\mathrm{T}}]E[\bar{W}Y]\} \end{aligned}$$

which achieves its minimum at $\theta_0 = (\beta_0, \gamma_0)$. Thus we obtain the consistency of $\hat{\theta}_0$.

Next consider the asymptotic presentation of $\hat{\theta}_0 - \theta_0$. the The estimator $\hat{\theta}_0$ satisfies the first order condition: $\partial Q_n(\hat{\theta}_0)/\partial\theta = 0$. By Taylor expansion and the mean value theorem, $\hat{\theta}_0 - \theta_0$ can be decomposed into

$$\begin{aligned} &\Big\{\Big[\frac{\partial^2 g^{\mathrm{T}}(\tilde{\boldsymbol{X}}\bar{\beta},\bar{\gamma})}{\partial\theta\partial\theta^{\mathrm{T}}}\tilde{\boldsymbol{D}}\Big](\tilde{\boldsymbol{D}}^{\mathrm{T}}\tilde{\boldsymbol{D}})^{-1}\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{Y} - \boldsymbol{D}(\tilde{\boldsymbol{D}}^{\mathrm{T}}\tilde{\boldsymbol{D}})^{-1}\tilde{\boldsymbol{D}}^{\mathrm{T}}g(\tilde{\boldsymbol{X}}\bar{\beta},\bar{\gamma})) \\ &-\Big[\frac{\partial g^{\mathrm{T}}(\tilde{\boldsymbol{X}}\bar{\beta},\bar{\gamma})}{\partial\theta}\tilde{\boldsymbol{D}}\Big](\tilde{\boldsymbol{D}}^{\mathrm{T}}\tilde{\boldsymbol{D}})^{-1}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D})(\tilde{\boldsymbol{D}}^{\mathrm{T}}\tilde{\boldsymbol{D}})^{-1}[\frac{\partial g^{\mathrm{T}}(\tilde{\boldsymbol{X}}\bar{\beta},\bar{\gamma})}{\partial\theta}\tilde{\boldsymbol{D}}]\Big\}^{-1} \\ &\times\Big[\frac{\partial g^{\mathrm{T}}(\tilde{\boldsymbol{X}}\beta_0,\gamma_0)}{\partial\theta}\tilde{\boldsymbol{D}}\Big](\tilde{\boldsymbol{D}}^{\mathrm{T}}\tilde{\boldsymbol{D}})^{-1}\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{Y} - \boldsymbol{D}(\tilde{\boldsymbol{D}}^{\mathrm{T}}\tilde{\boldsymbol{D}})^{-1}\tilde{\boldsymbol{D}}^{\mathrm{T}}g(\tilde{\boldsymbol{X}}\beta_0,\gamma_0)) \end{aligned}$$

where $\bar{\theta} = (\bar{\beta}; \bar{\gamma})$ is a vector satisfying $\|\bar{\theta} - \theta\| \leq \|\hat{\theta}_0 - \theta_0\|$. By the LLNs,

$$\frac{1}{N}\frac{\partial g^{\mathrm{T}}(\tilde{\boldsymbol{X}}\beta, \gamma)}{\partial \theta}\tilde{\boldsymbol{D}} = \frac{1}{N}\sum_{s=1}^{N}\frac{\partial g(\beta^{\mathrm{T}}\tilde{x}_s, \gamma)}{\partial \theta}\bar{w}_s^{\mathrm{T}} \rightarrow_p E[\frac{\partial g(\beta^{\mathrm{T}}X, \gamma)}{\partial \theta}\bar{W}^{\mathrm{T}}],$$
$$\frac{1}{N}\frac{\partial^2 g^{\mathrm{T}}(\tilde{\boldsymbol{X}}\beta, \gamma)}{\partial \theta \partial \theta^{\mathrm{T}}}\tilde{\boldsymbol{D}} \rightarrow_p E[\frac{\partial^2 g(\beta^{\mathrm{T}}X, \gamma)}{\partial \theta \partial \theta^{\mathrm{T}}}\bar{W}].$$

On the other hand,

$$\begin{aligned}
& \frac{1}{n}\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{Y} - \boldsymbol{D}(\tilde{\boldsymbol{D}}^{\mathrm{T}}\tilde{\boldsymbol{D}})^{-1}\tilde{\boldsymbol{D}}^{\mathrm{T}}g(\tilde{\boldsymbol{X}}\beta, \gamma)) \\
= \; & \frac{C_n}{n}\sum_{i=1}^{n}\bar{w}_i G(x_i) + \frac{1}{n}\sum_{i=1}^{n}\bar{w}_i\varepsilon_i + \frac{1}{n}\sum_{i=1}^{n}\bar{w}_i(g(\beta^{\mathrm{T}}x_i, \gamma) \\
& -E^{-1}[\bar{W}\bar{W}^{\mathrm{T}}]E[\bar{W}^{\mathrm{T}}g(\beta^{\mathrm{T}}X, \gamma)]\bar{w}_i) \\
& -\left(\frac{1}{n}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}\right)\left[\frac{1}{N}\tilde{\boldsymbol{D}}^{\mathrm{T}}\tilde{\boldsymbol{D}}\right]^{-1}\frac{1}{N}\sum_{s=1}^{N}\bar{\tilde{w}}_s(g(\beta^{\mathrm{T}}\tilde{x}_s, \gamma) \\
& -E^{-1}[\bar{W}\bar{W}^{\mathrm{T}}]E[\bar{W}^{\mathrm{T}}g(\beta^{\mathrm{T}}X, \gamma)]\bar{\tilde{w}}_s) \\
= \; & C_n E[\bar{W}G(x)] + O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{1}{\sqrt{N}}).
\end{aligned}$$

Hence

$$\begin{aligned}
\hat{\theta}_0 - \theta_0 \;=\; & \left\{E[\frac{\partial g(\beta_0^{\mathrm{T}}X, \gamma_0)}{\partial \theta}\bar{W}^{\mathrm{T}}]E^{-1}[\bar{W}\bar{W}^{\mathrm{T}}]E[\frac{\partial g(\beta_0^{\mathrm{T}}X, \gamma_0)}{\partial \theta^{\mathrm{T}}}\bar{W}]\right\}^{-1} \\
& \times\left\{E[\frac{\partial g(\beta_0^{\mathrm{T}}X, \gamma_0)}{\partial \theta}\bar{W}^{\mathrm{T}}]E^{-1}[\bar{W}\bar{W}^{\mathrm{T}}]\right\} \\
& \times\left(C_n E[\bar{W}G(x)] + O_p(\frac{1}{\sqrt{n}}) + O_p(\frac{1}{\sqrt{N}})\right).
\end{aligned}$$

This completes the proof of part (2) of Proposition 2. □

S3 Proof of Theorem 4

Denote $\zeta = \text{Cov}(X, W)\Sigma_W^{-1}W$. In the discretization step, we construct new samples $(\zeta_i, I(y_i \le t))$. For each t, we estimate $\Lambda(t)$ which spans $S_{I(Y\le t)|\zeta}$ by using SIR and denote the estimate by $\Lambda_n(t)$. In the expectation step, we estimate $\Lambda = E[\Lambda(t)]$, which spans $S_{Y|\zeta}$, by $\Lambda_{n,n} = n^{-1}\sum_{j=1}^n \Lambda_n(y_j)$. Let $\lambda_1 > \lambda_2 > \cdots > \lambda_q > \lambda_{q+1} = 0 = \cdots = \lambda_p$ be the descending sequence of eigenvalues of the matrix Λ and $\hat\lambda_1 \ge \hat\lambda_2 \ge \cdots \ge \hat\lambda_p$ be the descending sequence of eigenvalues of the matrix $\Lambda_{n,n}$. Recall that the D_n was selected as $\sqrt{n}$. Define the objective function as

$$G(l) = \frac{n}{2} \times \frac{\sum_{i=1}^l \{\log(\hat\lambda_i + 1) - \hat\lambda_i\}}{\sum_{i=1}^p \{\log(\hat\lambda_i + 1) - \hat\lambda_i\}} - 2 \times n^{1/2} \times \frac{l(l+1)}{2p}.$$

We shall prove that for any $l > 1$, $P(G(1) > G(l)) \to 1$, i.e., $P(\hat q = 1) \to 1$.

$$G(1) - G(l) = n^{1/2} \times \frac{l(l+1) - 2}{p} - \frac{n}{2} \times \frac{\sum_{i=2}^l \{\log(\hat\lambda_i + 1) - \hat\lambda_i\}}{\sum_{i=1}^p \{\log(\hat\lambda_i + 1) - \hat\lambda_i\}}$$

If $\Lambda_{n,n} - \Lambda = O_p(C_n)$, then $\hat\lambda_i - \lambda_i = O_p(C_n)$. By the second order Taylor Expansion, we have $\log(\hat\lambda_i+1) - \hat\lambda_i = -\hat\lambda_i^2 + o_p(\hat\lambda_i^2)$. Thus, $\sum_{i=2}^l \{\log(\hat\lambda_i+1) - \hat\lambda_i\} = O_p(C_n^2)$ and $\sum_{i=1}^p \{\log(\hat\lambda_i + 1) - \hat\lambda_i\}$ converge to a negative constant in probability. Since $nC_n^2/n^{1/2} \to 0$ and $l(l+1) > 2$, $P(G(1) > G(l)) \to 1$.

Now we check the condition of $\Lambda_{n,n} - \Lambda = O_p(C_n)$. First, we investigate the convergence rate of $\Lambda_n(t) - \Lambda(t)$ for any fixed t. We have

$$\Lambda(t) = \Sigma_\zeta^{-1}\text{Var}(E[\zeta|\tilde Y(t)])p(1-p) = \Sigma_X^{-1}\Sigma_W\Sigma_X^{-1}\text{Var}(E[\zeta|\tilde Y(t)])p(1-p).$$

It is easy to see that

$$\mathrm{Var}(E[\zeta|\tilde{Y}(t)]) = (u_1 - u_0)(u_1 - u_0)^{\mathrm{T}} p(1-p)$$

where $p = P(Y \le t) = E(I(Y \le t))$, $u_i = E[\zeta|\tilde{Y}(t) = i]$, $i = 0, 1$. Further, $u_1 - u_0$ can be rewritten as

$$u_1 - u_0 = \{E[\zeta I(Y \le t)] - E[\zeta]E[I(Y \le t)]\}/(p(1-p)).$$

We can use the matrix

$$\Lambda(t) = \Sigma_X^{-1}\Sigma_W\Sigma_X^{-1}\left[E\{(\zeta - E(\zeta))I(Y \le t)\}\right]\left[E\{(\zeta - E(\zeta))I(Y \le t)\}\right]^{\mathrm{T}}$$

to identify the central subspace we want. Denote $m(t) = E[(\zeta - E(\zeta))I(Y \le t)]$. The sample version of $m(t)$ is $\hat{m}(t) = \frac{1}{n}\sum_{i=1}^n (\zeta_i - \bar{\zeta})I(y_i \le t)$, where $\zeta_i = \hat{\mathrm{Cov}}(X, W)\hat{\Sigma}_W^{-1} w_i$ and $\bar{\zeta} = (1/n)\sum_{i=1}^n \zeta_i$. Let Y_a be the response under the local alternative, then

$$\begin{aligned}
\hat{m}(t) - m(t) &= \frac{1}{n}\sum_{i=1}^n (\zeta_i - \bar{\zeta})I(y_i \le t) - E\{(\zeta - E(\zeta))I(Y \le t)\} \\
&= \frac{1}{n}\sum_{i=1}^n (\zeta_i - \bar{\zeta})I(y_i \le t) - E\{(\zeta - E(\zeta))I(Y_a \le t)\} \\
&\quad + E\{(\zeta - E(\zeta))I(Y_a \le t)\} - E\{(\zeta - E(\zeta))I(Y \le t)\}.
\end{aligned}$$

The convergence rate of the first term in the right hand side is $O_p(\sqrt{n})$. For simplicity, we assume $E(\zeta) = 0$. The second term is

$$E[\zeta I(Y_a \le t)] - E[\zeta I(Y \le t)] = E\left\{\zeta[P(Y_a \le t|\zeta) - P(Y \le t|\zeta)]\right\}$$

Since $\zeta = \Sigma_X \Sigma_W^{-1} W$,

$$
\begin{aligned}
& P(Y_a \le t|\zeta) - P(Y \le t|\zeta) = P(Y_a \le t|W) - P(Y \le t|W) \\
= \;& F_{Y|W}(t - C_n E[G(B^{\mathrm{T}}X)|B^{\mathrm{T}}W]) - F_{Y|W}(t) \\
= \;& -C_n E[G(B^{\mathrm{T}}X)|B^{\mathrm{T}}W] f_{Y|W}(t) + O_p(C_n^2).
\end{aligned}
$$

Thus, we have $E\{(\zeta - E(\zeta))I(Y_a \le t)\} - E\{(\zeta - E(\zeta))I(Y \le t)\} = O_p(C_n)$. Altogether, $\Lambda_n(t) - \Lambda(t) = O_p(C_n)$, for each $t \in \mathbb{R}$. Finally, similar to the proof for Theorem 3.2 of Li et al. (2008) the condition $\Lambda_{n,n} - \Lambda = O_p(C_n)$ holds.

□

S4 Proof of Theorem 5

In this subsection, we first prove (ii) which is the large sample property of V_n under the local alternatives and then give a sketch of the proof of (i). For the local alternatives in (3.5), according to Theorem 4, $\hat{q} = 1$ with a probability going to 1. Thus, we can only work on the event that $\hat{q} = 1$. Note that $\hat{B}(\hat{q})$ converges to $B_0 = \pm\beta_0/\|\beta_0\|$ in probability rather than the $p \times q$ matrix B that is the dimension reduction base matrix of the central mean subspace. Recall the notations

$$\eta = g(\beta_0^{\mathrm{T}} X, \gamma_0) - r(b_0^{\mathrm{T}} W, \theta_0), \quad \xi^2(Z) = E[\eta^2|Z],$$
$$\Delta(Z) = E[G(B^{\mathrm{T}} X)|Z]. \tag{S4.1}$$

The variance of ε is σ^2. Write Z as $\tilde{Z}$, when W is replaced by validation data $\tilde{W}$. Note $Z = B_0^{\mathrm{T}} W = \pm b_0^{\mathrm{T}} W$. The proof for the case $B_0 = -b_0$ is similar to that for the case $B_0 = b_0$. Also in practise, we can change the sign of $\hat{B}$ to make sure $B_0 = b_0$. So, in the following proof, we only discuss the case $B_0 = b_0$. To proceed further, we need some more notation as follows:

$$z_i = B^{\mathrm{T}} w_i, \quad g_i = g(\beta_0^{\mathrm{T}} x_i, \gamma_0), \quad r_i = r(b^{\mathrm{T}} w_i, \theta),$$
$$\eta_i = g_i - r_i, \quad \Delta_i = \Delta(z_i). \tag{S4.2}$$

Write $\tilde{z}_s$, $\tilde{g}_s$, $\tilde{r}_s$ and $\tilde{\eta}_s$ for the entities in (S4.2) when w_i is replaced by validation data $\tilde{w}_s$ in there. When θ_0 and B_0 are respectively replaced by their estimators $\hat{\theta}_0$ and $\hat{B}(\hat{q})$ in the above definitions, write the respective $\hat{z}_i$, $\hat{g}_i$, $\hat{r}_i$ and $\hat{\eta}_i$ for z_i, g_i, r_i and η_i, and similarly write the respective $\hat{\tilde{z}}_s$, $\hat{\tilde{g}}_s$, $\hat{\tilde{r}}_s$ and $\hat{\tilde{\eta}}_s$ for $\tilde{z}_s$, $\tilde{g}_s$, $\tilde{r}_s$ and $\tilde{\eta}_s$. In addition, let $G_i = G(z_i)$, where G is in (3.5). Plug $y_i = g_i + C_n G_i + \varepsilon_i$ into V_n, we obtain that $V_n = V_{n1} + V_{n2} + V_{n3} + V_{n4}$, where

$$\begin{aligned}
V_{n1} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(\hat{z}_i-\hat{z}_j)(e_i+C_nG_i)(e_j+C_nG_j),\\
V_{n2} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(\hat{z}_i-\hat{z}_j)(e_i+C_nG_i)(r_j-\hat{r}_{j(2)}),\\
V_{n3} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(\hat{z}_i-\hat{z}_j)(r_i-\hat{r}_{i(1)})(e_j+C_nG_j),\\
V_{n4} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(\hat{z}_i-\hat{z}_j)(r_i-\hat{r}_{i(1)})(r_j-\hat{r}_{j(2)}).
\end{aligned}$$

We now deal with V_{ni}'s in the following steps.

Step S4.1. $nh^{1/2}V_{n1} \to_D N(\nu_1,\tau_1)$, where

$$\nu_1 = E[\Delta(Z)^2 f_Z(Z)], \quad \tau_1 = 2\int K^2(u)du\int(\sigma^2+\xi^2(z))^2 f_Z^2(z)dz.$$

Proof: Write V_{n1} as $I_1 + 2C_nI_2 + C_n^2 I_3$ where

$$\begin{aligned}
I_1 &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(\hat{z}_i-\hat{z}_j)e_ie_j,\\
I_2 &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(\hat{z}_i-\hat{z}_j)e_iG_j,\\
I_3 &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(\hat{z}_i-\hat{z}_j)G_iG_j.
\end{aligned}$$

Rewrite $I_1 = I_{1,1} + I_{1,2}$, where

$$\begin{aligned}
I_{1,1} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i-z_j)e_ie_j,\\
I_{1,2} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} (K_h(\hat{z}_i-\hat{z}_j)-K_h(z_i-z_j))e_ie_j.
\end{aligned}$$

Note $e_i = y_i - r_i = \varepsilon_i + \eta_i$ where η_i is in (S4.2). Thus $E[e_i^2|Z_i] = \sigma^2 + \xi^2(Z_i)$. Following Lemma 3.3a of Zheng (1996), we obtain $nh^{1/2}I_{1,1} \to_D N(0, \tau_1)$, where

$$\tau_1 = 2\int(\sigma^2 + \xi^2(z))^2 f_Z^2(z)dz \int K^2(u)du.$$

The Taylor expansion yields that $I_{1,2} = I_{1,2}^*(1 + o_p(1))$ where

$$\begin{aligned} I_{1,2}^* &= \frac{(\hat{B} - B_0)^{\mathrm{T}}}{h}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K'(\frac{z_i - z_j}{h})\frac{w_i - w_j}{h}e_i e_j \\ &\quad \times I(|z_i - z_j| \leq h \text{ or } |\hat{z}_i - \hat{z}_j| \leq h). \end{aligned}$$

Note

$$|(\hat{z}_i - \hat{z}_j) - (z_i - z_j)| \leq \|\hat{B} - B_0\| \max_{i,j} \|w_i - w_j\| = O_p(C_n \log n).$$

When n is large enough, $|(\hat{z}_i - \hat{z}_j) - (z_i - z_j)| \ll h$. Then we have, for large n,

$$I(|z_i - z_j| \leq h \text{ or } |\hat{z}_i - \hat{z}_j| \leq h) \leq I(|z_i - z_j| \leq 2h).$$

Thus $E[(I_{1,2}^*)^2]$ is bounded above by

$$\begin{aligned} &\frac{1}{n^2h^4}\|\hat{B} - B_0\|^2\|K'\|_\infty^2 \max_{i,j}\|w_i - w_j\|^2 E\left[e_i^2 e_j^2 I(|z_i - z_j| \leq 2h)\right] \\ &= O_p(\frac{\log^2 n}{n^3 h^{7/2}}) = o_p(1), \end{aligned}$$

where K' denote the first order derivative of the kernel function K and $\|.\|_\infty$ is the uniform norm. Then $E[n^2hI_{1,2}^2] = O(\log^2 n/(nh^{5/2})) = o(1)$. By Chebyshev's inequality, we obtain $nh^{1/2}I_{1,2}$ is asymptotical negligible.

Next, consider I_2. Rewrite $I_2 = I_{2,1} + I_{2,2}$, where

$$\begin{aligned} I_{2,1} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} K_h(z_i - z_j) e_i G_j, \\ I_{2,2} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} (K_h(\hat{z}_i - \hat{z}_j) - K_h(z_i - z_j)) e_i G_j. \end{aligned}$$

To compute the first two moments, $E[I_{2,1}] = 0$ and $E[n^2 h C_n^2 I_{2,1}^2] = O(h^{1/2})$. Thus, by Chebyshev's inequality, $nh^{1/2} C_n I_{2,1} = o_p(1)$. As to $I_{2,2}$, by Taylor expansion, $I_{2,2} = I_{2,2}^*(1 + o_p(1))$ where

$$\begin{aligned} I_{2,2}^* &= \frac{\hat{B} - B}{h} \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} K'(\frac{z_i - z_j}{h}) \frac{w_i - w_j}{h} e_i G_j \\ &\quad \times I(|z_i - z_j| \leq h \text{ or } |\hat{z}_i - \hat{z}_j| \leq h). \end{aligned}$$

Similar to $I_{1,2}$, we obtain that $E[n^2 h C_n^2 I_{2,2}] \leq O(\log^2 n / (nh^2))$. Then, by Chebyshev's inequality, $nh^{1/2} C_n I_{2,2} = o_p(1)$. Combining the results of $I_{2,1}$ and $I_{2,2}$, we know that $nh^{1/2} C_n I_2 = o_p(1)$.

To finish the proof of this step, it suffices to show $I_3 \to_p \nu_1$. Write I_3 as $I_{3,1} + I_{3,2}$ where

$$\begin{aligned} I_{3,1} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} K_h(z_i - z_j) G_i G_j, \\ I_{3,2} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} (K_h(\hat{z}_i - \hat{z}_j) - K_h(z_i - z_j)) G_i G_j. \end{aligned}$$

By the Law of Large Numbers, $I_{3,1} \to_p \nu_1$. In addition, by Taylor expansion and the fact $\hat{B} \to_p B_0$, it is easy to see $I_{3,2} = o_p(1)$.

Hence the proof of Step S4.1 is finished. $\square$

STEP S4.2. $nh^{1/2}V_{n2} \to_D N(\nu_2, 2\lambda^{-1}\tau_2)$, where

$$\begin{aligned} \nu_2 &= -E\{\Delta(Z)E[\frac{\partial g(\beta_0^{\mathrm{T}}\tilde{x}_s, \gamma_0)}{\partial\theta}|Z]f_Z(Z)\}H(\theta_0), \\ \tau_2 &= \int K^2(u)du \int (\sigma^2 + \xi^2(z))\xi^2(z)f_Z^2(z)dz, \end{aligned} \tag{S4.3}$$

and $H(\theta_0)$ is defined in Proposition 2.

Proof: Rewrite V_{n2} as

$$\begin{aligned} V_{n2} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(\hat{z}_i - \hat{z}_j)e_i(r_j - \hat{r}_{j(2)}) \quad \text{(S4.4)} \\ &\quad + \frac{C_n}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(\hat{z}_i - \hat{z}_j)G_i(r_j - \hat{r}_{j(2)}) \\ &=: V_{n2,1} + C_n V_{n2,2}, \qquad \text{say.} \end{aligned}$$

First, deal with the term $V_{n2,1}$. It can be decomposed as

$$\begin{aligned} V_{n2,1} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i - z_j)e_i(r_j - \hat{r}_{j(2)}) \\ &\quad + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} (K_h(\hat{z}_i - \hat{z}_j) - K_h(z_i - z_j))e_i(r_j - \hat{r}_{j(2)}). \end{aligned}$$

Recalling the definition of the estimator of $r_{(2)}(b_0^{\mathrm{T}}w, \theta_0)$ in (2.2), we have

$$\begin{aligned} r_j - \hat{r}_{j(2)} &= \frac{2}{N}\sum_{s=N/2+1}^{N} M_v(\hat{b}_0^{\mathrm{T}}w_j - \hat{b}_0^{\mathrm{T}}\tilde{w}_s)(r_j - \hat{\tilde{g}}_s) \\ &\quad \times (\frac{2}{N}\sum_{s=N/2+1}^{N} M_v(\hat{b}_0^{\mathrm{T}}w_j - \hat{b}_0^{\mathrm{T}}\tilde{w}_s))^{-1}, \end{aligned} \tag{S4.5}$$

where $\hat{\tilde{g}}_s$ is defined in (S4.2). In order to analyze $r_j - \hat{r}_{j(2)}$ further, we need the following entities. Let

$$\bar{f}_{N(2)}(x,b) = \frac{2}{N}\sum_{s=N/2+1}^{N} M_v(x - b^{\mathrm{T}}\tilde{w}_s), \tag{S4.6}$$

$$\begin{aligned} Q_{1(2)}(x,b) &= \frac{2}{N}\sum_{s=N/2+1}^{N} M_v(x - b^{\mathrm{T}}\tilde{w}_s)(r_j - \tilde{r}_s), \qquad (\mathrm{S4.7}) \\ Q_{2(2)}(x,b) &= \frac{2}{N}\sum_{s=N/2+1}^{N} M_v(x - b^{\mathrm{T}}\tilde{w}_s)(\tilde{r}_s - \tilde{g}_s), \\ Q_{3(2)}(x,b) &= \frac{2}{N}\sum_{s=N/2+1}^{N} M_v(x - b^{\mathrm{T}}\tilde{w}_s)(\tilde{g}_s - \hat{\tilde{g}}_s). \end{aligned}$$

Note $z_i = B_0^{\mathrm{T}} w_i = b_0^{\mathrm{T}} w_i$. The kernel function $M_v(\hat{b}_0^{\mathrm{T}} w_j - \hat{b}_0^{\mathrm{T}}\tilde{w}_s)$ in the numerator of (S4.5) can be rewritten as

$$M_v(z_j - \tilde{z}_s) + [M_v(\hat{b}_0^{\mathrm{T}} w_j - \hat{b}_0^{\mathrm{T}}\tilde{w}_s) - M_v(z_j - \tilde{z}_s)],$$

and the denominator can be decomposed as

$$\frac{1}{\bar{f}_{N(2)}(z_j, b_0)} + [\frac{1}{\bar{f}_{N(2)}(\hat{b}_0^{\mathrm{T}} w_j, \hat{b}_0)} - \frac{1}{\bar{f}_{N(2)}(z_j, b_0)}].$$

Further, write

$$r_j - \hat{\tilde{g}}_s = [r_j - \tilde{r}_s] + [\tilde{r}_s - \tilde{g}_s] + [\tilde{g}_s - \hat{\tilde{g}}_s].$$

Combining the above decompositions into (S4.5), $r_j - \hat{r}_{j(2)}$ can be decomposed into 12 terms, and then $V_{n2,1}$ can be decomposed into 24 terms. We only consider the following three terms that make non-negligible contribution. The remaining terms can be shown to be asymptotically negligible, in probability. Accordingly, consider

$$\begin{aligned} I_4 &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i-z_j)e_i Q_{1(2)}(z_j,b_0)\bar{f}_{N(2)}^{-1}(z_j,b_0), \\ I_5 &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i-z_j)e_i Q_{2(2)}(z_j,b_0)\bar{f}_{N(2)}^{-1}(z_j,b_0), \\ I_6 &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i-z_j)e_i Q_{3(2)}(z_j,b_0)\bar{f}_{N(2)}^{-1}(z_j,b_0) \end{aligned}$$

where $\bar{f}_{N(2)}$ is defined in (S4.6), and $Q_{1(2)}$, $Q_{2(2)}$, $Q_{3(2)}$ are in (S4.7).

We first prove that $nh^{1/2}I_4 = o_p(1)$. Rewrite $I_4 = n^{-1}\sum_{j=1}^{n} I_{41}(z_j) \times I_{42}(z_j,b_0)$, where

$$I_{41}(z_j) = \frac{1}{(n-1)}\sum_{i\neq j}^{n} K_h(z_i-z_j)e_i, \;\; I_{42}(z_j,b_0) = \frac{Q_{1(2)}(z_j,b_0)}{\bar{f}_{N(2)}(z_j,b_0)}.$$

Thus, the application of Cauchy - Schwarz inequality yields that $|I_4| \leq \sqrt{(1/n)\sum_{j=1}^{n} I_{41}^2(z_j)} \times \sqrt{(1/n)\sum_{j=1}^{n} I_{42}^2(z_j,b_0)}$. We only need to bound the conditional expectations $E[I_{41}^2(z_j)]$ and $E[I_{42}^2(z_j,b_0)]$ when z_j is given. For $I_{41}(z_j)$,

$$\begin{aligned} E[I_{41}^2(z_j)] &= \frac{1}{(n-1)^2}E[(\sum_{i\neq j}^{n} K_h(z_i-z_j)e_i)^2] \\ &= \frac{1}{(n-1)h^2}E[K^2(\frac{z_i-z_j}{h})e_i^2] = O(\frac{1}{nh}). \end{aligned}$$

For I_{42}, we can obtain that given z_j,

$$|I_{42}(z_j,b_0)| \leq \left|\frac{Q_{1(2)}(z_j,b_0)}{f_Z(z_j)}\right| \sup_{z_j}\left|\frac{f_Z(z_j)}{\bar{f}_{N(2)}(z_j,b_0)}\right|,$$

where f_Z is the density of Z. Since

$$\sup_{z_j} |\bar{f}_{N(2)}(z_j, b_0) - f_Z(z_j)| = o_p(1), \ \ \sup_{z_j} \left| \frac{\bar{f}_{N(2)}(z_j, b_0)}{f_Z(z_j)} - 1 \right| = o_p(1),$$

and f_Z is uniformly bounded below, we only need to bound $Q^2_{1(2)}(z_j, b_0)$ in the numerators. By Conditions (f),(r) and (M),

$$\begin{aligned} E[Q^2_{1(2)}(z_j, b_0)] &= \frac{N(N-2)}{N^2 v^2} E[M(\frac{z_j - \tilde{z}_s}{v})(r_j - \tilde{r}_s) M(\frac{z_j - \tilde{z}_{s'}}{v})(r_j - \tilde{r}_{s'})] \\ &\quad + \frac{2}{N v^2} E[M^2(\frac{z_j - \tilde{z}_s}{v})(r_j - \tilde{r}_s)^2] \\ &\leq C_1 v^4 + N^{-1} C_2 v, \end{aligned}$$

where C_1 and C_2 are two constants. Thus $E[I^2_{42}(z_j, b_0)]$ is bounded above by $C_1 v^4 + C_2 v/N$, in probability. Summarizing the results of $E[I^2_{41}]$ and $E[I^2_{42}]$, we have $E[n^2 h I_4^2] \leq n h^{1/2} O_p(\frac{1}{nh}(v^4 + \frac{v}{N})) = o_p(1)$.

Consider I_5. Rewrite it as $I_5 = I_{51} + I_{52}$, where

$$I_{51} = E[I_5 | \tilde{\eta}_s, \tilde{z}_s, z_i, e_i], \ \ I_{52} = (I_5 - E[I_5 | \tilde{\eta}_s, \tilde{z}_s, z_i, e_i]).$$

Note $z_j = B_0^{\mathrm{T}} w_j = b_0^{\mathrm{T}} w_j$. Thus,

$$\begin{aligned} I_{51} &= \frac{2}{nN} \sum_{i=1}^{n} \sum_{s=N/2+1}^{N} e_i \tilde{\eta}_s \int \frac{1}{h} K(\frac{z_i - z_j}{h}) \frac{1}{v} M(\frac{z_j - \tilde{z}_s}{v}) dz_j \\ &= \frac{2}{nN} \sum_{i=1}^{n} \sum_{s=N/2+1}^{N} e_i \tilde{\eta}_s \int \frac{1}{h} K(\frac{z_i - \tilde{z}_s - uv}{h}) \frac{1}{v} M(u) d(\tilde{z}_s + uv). \end{aligned}$$

Further,

$$\int \frac{1}{h} K(\frac{z_i - \tilde{z}_s - uv}{h}) M(u) du = \frac{1}{h} K(\frac{z_i - \tilde{z}_s}{h}) + \frac{1}{h} K''(\frac{z_i - \tilde{z}_s}{h}) \frac{u^2 v^2}{h^2}.$$

Thus, $I_{51} = \frac{2}{nN}\sum_{i=1}^{n}\sum_{s=N/2+1}^{N} e_i\tilde{\eta}_s K_h(z_i - \tilde{z}_s)(1 + o_p(1))$. By Central Limit Theorem we have

$$\sqrt{\frac{nN}{2}}h^{1/2}I_{5,1} \to_D N(0, \int K^2(u)du \int (\sigma^2 + \xi^2(z))\xi^2(z)f_Z^2(z)dz),$$

where σ^2 is the variance of ε and $\xi^2(Z)$ is defined in (S4.1). By some elementary calculations, we can derive that $E[(I_{52})^2] = O_p(1/(n^2Nhv_N))$. Chebyshev's inequality yields that $nh^{1/2}I_{52} = o_p(1)$. Hence

$$nh^{1/2}I_5 \to_D N\Big(0, 2\lambda^{-1}\int K^2(u)du \int (\sigma^2 + \xi^2(z))\xi^2(z)f_Z^2(z)dz\Big).$$

Now consider I_6. Recall the definition of $Q_{3(2)}$ in (S4.7) and the definition of $\tilde{g}$ below (S4.2). Taylor expansion of the function $\tilde{g}$ yields that $I_6 = I_6^*(\theta_0 - \hat{\theta}_0)(1 + o_p(1))$, where

$$\begin{aligned} I_6^* &= \frac{2}{Nn(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{K_h(z_i - z_j)e_i}{\bar{f}_{N(2)}(z_j, b_0)}\sum_{s=N/2+1}^{N} M_v(z_j - \tilde{z}_s)\frac{\partial g(\beta_0^{\mathrm{T}}\tilde{x}_s, \gamma_0)}{\partial\theta} \\ &=: \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{i\neq j}^{n} K_h(z_i - z_j)e_i I_{62}(z_j, b_0), \qquad \text{say.} \end{aligned}$$

It is easy to see that for any given z_j, $I_{62}(z_j, b_0) \to_p E[\frac{\partial g(\beta_0^{\mathrm{T}}\tilde{x}, \gamma_0)}{\partial\theta}|z_j]$ by noticing that $\tilde{x}$ has the same distribution as that of x. By Lemma 2 of Guo et al. (2016),

$$\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{i\neq j}^{n} K_h(z_i - z_j)e_i E[\frac{\partial g(\beta_0^{\mathrm{T}}\tilde{x}, \gamma_0)}{\partial\theta}|z_j] = O_p(\frac{1}{\sqrt{n}}).$$

Similarly, as in the proof for I_4, we can also derive that as $N \to \infty$,

$$\sup_z |I_{62}(z, b_0) - E[\frac{\partial g(\beta_0^{\mathrm{T}}\tilde{x}, \gamma_0)}{\partial\theta}|z]| \leq O(v^2 + \log(N)/\sqrt{Nv})$$

and then

$$\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{i\neq j}^{n}K_h(z_i-z_j)e_i(I_{62}(z_j,b_0)-E[\frac{\partial g(\beta_0^{\mathrm{T}}\tilde{x},\gamma_0)}{\partial\theta}|z_j])=o_p(\frac{1}{\sqrt{n}}).$$

Hence $nh^{1/2}I_6=o_p(1)$. Combining the above results for I_4, I_5 and I_6 with the fact that the remaining 21 terms tend to zero, in probability, we obtain that $nh^{1/2}V_{n2,1}\to_D N(0,2\lambda^{-1}\tau_2)$, where τ_2 is in (S4.3).

Next, consider the second term $V_{n2,2}$ of the decomposition (S4.4). Rewrite

$$\begin{aligned}V_{n2,2} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}K_h(z_i-z_j)G_i(r_j-\hat{r}_{j(2)})\\ &\quad+\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(K_h(\hat{z}_i-\hat{z}_j)-K_h(z_i-z_j))G_i(r_j-\hat{r}_{j(2)}).\end{aligned}$$

Similarly as the decomposition in (S4.5), $V_{n2,2}$ can also be decomposed into 24 terms. Again, we only give the detail about how to treat the three leading terms. Again, the remaining 21 terms tend to zero, in probability. The three leading terms are:

$$\begin{aligned}I_7 &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}K_h(z_i-z_j)G_iQ_{1(2)}(z_j,b_0)/\bar{f}_{N(2)}(z_j,b_0),\\ I_8 &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}K_h(z_i-z_j)G_iQ_{2(2)}(z_j,b_0)/\bar{f}_{N(2)}(z_j,b_0),\\ I_9 &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}K_h(z_i-z_j)G_iQ_{3(2)}(z_j,b_0)/\bar{f}_{N(2)}(z_j,b_0),\end{aligned}$$

where $Q_{1(2)}$, $Q_{2(2)}$, $Q_{3(2)}$ and $\bar{f}_{N(2)}$ are defined in (S4.7) and (S4.6). Recall that $C_n=n^{-1/2}h^{-1/4}$ and $E[Q_{1(2)}^2(z_j,b_0)]\leq C_1v^4+C_2v/N$ given z_j, which was proved when we handled I_4. By the Cauchy–Schwarz inequality,

$$|nh^{1/2}C_nI_7| \le O_p\Big(n^{1/2}h^{1/4}\sqrt{C_1v^4+C_2v/N}\Big) = o_p(1).$$

To deal with I_8, decompose $I_8 = I_{81} + I_{82}$, with

$$I_{81} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i-z_j)G_iQ_{2(2)}(z_j,b_0)/f_Z(z_j),$$
$$I_{82} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i-z_j)G_iQ_{2(2)}(z_j,b_0)[\frac{1}{\bar{f}_{N(2)}(z_j,b_0)} - \frac{1}{f_Z(z_j)}],$$

where f_Z is the density of Z. By some elementary calculations, one can verify that $E[I_{81}^2] = O_p(1/N)$. This implies $nh^{1/2}C_nI_{81} = o_p(1)$. Next, consider I_{82}. By the Cauchy–Schwarz inequality, I_{82}^2 is bounded above by a product of $\sum_{j=1}^n I_{821}^2(z_j)/n$ and $\sum_{j=1}^n I_{822}^2(z_j)/n$, where

$$I_{821}(z_j) = \frac{1}{n}\sum_{i\neq j} K_h(z_i-z_j)G_i,$$
$$I_{822}(z_j) = Q_{2(2)}(z_j,b_0)[\frac{1}{\bar{f}_{N(2)}(z_j,b_0)} - \frac{1}{f_Z(z_j)}].$$

Now we bound $E[I_{821}^2(z_j)]$ and $E[I_{822}^2(z_j)]$. Clearly, conditional on z_j, $E[I_{821}^2(z_j)] = O(1)$, which in turn implies that $E\big\{\sum_{j=1}^n I_{821}^2(z_j)/n\big\} = O(1)$.

Next, note that

$$\frac{1}{n}\sum_{j=1}^{n} I_{822}^2(z_j) \le \frac{1}{n}\sum_{j=1}^{n} Q_{2(2)}^2(z_j,b_0)\sup_z|\frac{1}{\bar{f}_{N(2)}(z,b_0)} - \frac{1}{f_Z(z)}|^2$$
$$\le O_p(v^2+\log N/\sqrt{Nv})\frac{1}{n}\sum_{j=1}^{n} Q_{2(2)}^2(z_j,b_0).$$

The second inequality is from the fact that f_Z is bounded below and $\sup_z|\bar{f}_{N(2)}(z,b_0) - f_Z(z)| = O_p(v^2+\log N/\sqrt{Nv})$. By $E[(\tilde{r}_s-\tilde{g}_s)|\tilde{z}_s] = 0$,

$E[Q^2_{2(2)}(z_j, b_0)] \le O(1/(Nv))$ given z_j which implies

$$E\Big\{\sum_{j=1}^{n} Q^2_{2(2)}(z_j, b_0)/n\Big\} \le O(1/(Nv)).$$

Thus $\sum_{j=1}^{n} I^2_{822}(z_j)/n$ is bounded above by $O_p(1/Nv)O_p(v^2 + \log N/\sqrt{Nv}) = o_p(1/(nh^{1/2}C_n)^2)$. Combining these results, we obtain that

$$\left|nh^{1/2}C_n I_{82}\right| \le nh^{1/2}C_n o_p(1/(nh^{1/2}C_n)) = o_p(1).$$

The above results about I_{81} and I_{82} in turn yield that $nh^{1/2}C_n I_8 = o_p(1)$.

Now we analyze I_9. Recall the definitions that $G_i = G(B^{\mathrm{T}}x_i)$ and $\Delta_i = E[G(B^{\mathrm{T}}X)|Z = z_i]$. Write $I_9 = I_{91} + I_{92}$, where

$$I_{91} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i - z_j)\Delta_i Q_{3(2)}(z_j, b_0)/\bar{f}_{N(2)}(z_j, b_0)$$
$$I_{92} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i - z_j)(G_i - \Delta_i)Q_{3(2)}(z_j, b_0)/\bar{f}_{N(2)}(z_j, b_0).$$

For I_{92}, $E[G_i - \Delta_i|Z_i] = 0$. Thus, $nh^{1/2}I_{92} = o_p(1)$, at the same rate as I_6. So $nh^{1/2}C_n I_{92} = o_p(1)$. Next, we deal with I_{91}. Similar to I_8, rewrite $I_{91} = I_{911} + I_{912}$, where

$$I_{911} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i - z_j)\Delta_i Q_{3(2)}(z_j, b_0)/f_Z(z_j),$$
$$I_{912} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i - z_j)\Delta_i Q_{3(2)}(z_j, b_0) \times\Big[\frac{1}{\bar{f}_{N(2)}(z_j, b_0)} - \frac{1}{f_Z(z_j, b_0)}\Big].$$

Similar to I_{82}, we have $nh^{1/2}I_{912} = o_p(1)$, because $E[Q^2_{3(2)}(z_j, b_0)] = O_p(C_n^2)$.

Next, consider I_{911}. Write $E[I_{911}|z_i, \tilde{z}_s, \tilde{x}_s] = I^*_{911}(1 + o_p(1))$ where

$$I^*_{911} = \frac{2}{nN}\sum_{i=1}^{n}\sum_{s=N/2+1}^{N} K_h(z_i - \tilde{z}_s)\Delta_i(\tilde{g}_s - \hat{\tilde{g}}_s).$$

By the first order Taylor expansion,

$$I^*_{911} = \frac{2}{nN}\sum_{i=1}^{n}\sum_{s=N/2+1}^{N} K_h(z_i - \tilde{z}_s)\Delta_i\frac{\partial g(\beta_0^{\mathrm{T}}\tilde{x}_s, \gamma_0)}{\partial\theta}(\theta_0 - \hat{\theta}_0)(1 + o_p(1))$$

Combining the result of Proposition 2(2),

$$nh^{1/2}C_n I^*_{911} \to_p \quad \nu_2 = -E\{\Delta(Z)E[\tfrac{\partial g(\beta_0^{\mathrm{T}}\tilde{x}_s,\gamma_0)}{\partial\theta}|Z]f_Z(Z)\}H(\theta_0).$$

By computing the second moment of $I_{911} - I^*_{911}$ and using the Chebyshev's inequality, one can verify $nh^{1/2}C_n(I_{911} - I^*_{911}) = o_p(1)$. Hence $nh^{1/2}C_n I_9 \to \nu_2$. These results about I_7, I_8 and I_9 imply that $nh^{1/2}C_n V_{n2,2} \to_p \nu_2$. Hence Step S4.2 is finished. □

Step S4.3. $nh^{1/2}V_{n3} \to_D N(\nu_2, 2\lambda^{-1}\tau_2)$, where ν_2 and τ_2 are as in (S4.3).

Proof: The proof is similar to that pertaining to V_{n2} in STEP S4.2. The only difference is that instead of the representation (S4.5) we now use

$$\begin{aligned} r_i - \hat{r}_{i(1)} &= \frac{2}{N}\sum_{t=1}^{N/2} M_v(\hat{b}_0^{\mathrm{T}}w_i - \hat{b}_0^{\mathrm{T}}\tilde{w}_t)(r_i - \hat{\tilde{g}}_t) \\ &\quad \times\frac{2}{N}\sum_{t=1}^{N/2} M_v(\hat{b}_0^{\mathrm{T}}w_i - \hat{b}_0^{\mathrm{T}}\tilde{w}_t). \end{aligned} \tag{S4.8}$$

Further the definitions in (S4.6) and (S4.7) are changed into

$$\bar{f}_{N(1)}(x, b) = \frac{2}{N}\sum_{t=1}^{N/2} M_v(x - b^{\mathrm{T}}\tilde{w}_t),$$

and

$$\begin{aligned}
Q_{1(1)}(z_i, b) &= \frac{2}{N}\sum_{t=1}^{N/2} M_v(b^{\mathrm{T}}w_i - b^{\mathrm{T}}\tilde{w}_t)(r_i - \tilde{r}_t), \\
Q_{2(1)}(z_i, b) &= \frac{2}{N}\sum_{t=1}^{N/2} M_v(b^{\mathrm{T}}w_i - b^{\mathrm{T}}\tilde{w}_t)(\tilde{r}_t - \tilde{g}_t), \\
Q_{3(1)}(z_i, b) &= \frac{2}{N}\sum_{t=1}^{N/2} M_v(b^{\mathrm{T}}w_i - b^{\mathrm{T}}\tilde{w}_t)(\tilde{g}_t - \hat{\tilde{g}}_t).
\end{aligned}$$

We omit the details here. □

STEP S4.4. $nh^{1/2}V_{n4} \to_D N(\nu_3, 2\lambda^{-2}\tau_3)$, where

$$\begin{aligned}
\nu_3 &= H^{\mathrm{T}}(\theta_0)E\{E[\frac{\partial g(\beta_0^{\mathrm{T}}\tilde{x}_s, \gamma_0)}{\partial\theta}|Z]E[\frac{\partial g(\beta_0^{\mathrm{T}}\tilde{x}_s, \gamma_0)}{\partial\theta^{\mathrm{T}}}|Z]f_Z(Z)\}H(\theta_0), \\
\tau_3 &= 2\int K^2(u)du\int(\xi^2(z))^2 f_Z^2(z)dz. \qquad \text{(S4.9)}
\end{aligned}$$

and $H(\theta_0)$ is defined in Proposition 2.

Proof: By the same decompositions in (S4.5) and (S4.8), V_{n4} can be decomposed to 9 dominant terms, and seven of those are of order $o_p(1/(nh^{1/2}))$. We investigate the other two terms as follows:

$$\begin{aligned}
I_{10} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i - z_j)Q_{2(1)}(z_i, b_0)Q_{2(2)}(z_j, b_0) \\
&\quad \times \bar{f}_{N(1)}^{-1}(z_i, b_0)\bar{f}_{N(2)}^{-1}(z_j, b_0), \\
I_{11} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i - z_j)Q_{3(1)}(z_i, b_0)Q_{3(2)}(z_j, b_0) \\
&\quad \times \bar{f}_{N(1)}^{-1}(z_i, b_0)\bar{f}_{N(2)}^{-1}(z_j, b_0).
\end{aligned}$$

Similar to the proof of I_5, we have $Nh^{1/2}I_{10} \to_D N(0, 2\tau_3)$, where τ_3 is defined in (S4.9). Similarly as I_{91}, I_{11} can be rewritten as

$$\begin{aligned} I_{11} &= \frac{4}{N^2}\sum_{t=1}^{N/2}\sum_{s=N/2+1}^{N} K_h(\tilde{z}_t - \tilde{z}_s)(\tilde{g}_s - \hat{\tilde{g}}_s)(\tilde{g}_t - \hat{\tilde{g}}_t)(1 + o_p(1)) \\ &= (\theta_0 - \hat{\theta}_0)^{\mathrm{T}}\left[\frac{4}{N^2}\sum_{s=1}^{N/2}\sum_{t=N/2+1}^{N} K_h(\tilde{z}_t - \tilde{z}_s)\frac{\partial g(\beta_0^{\mathrm{T}}\tilde{x}_s, \gamma_0)}{\partial \theta} g' \frac{\partial g(\beta_0^{\mathrm{T}}\tilde{x}_t, \gamma_0)}{\partial \theta^{\mathrm{T}}}\right] \\ &\quad \times (\theta_0 - \hat{\theta}_0). \end{aligned}$$

By the Law of Large Numbers and Proposition 2, $nh^{1/2}I_{11}$ converges to ν_3 in probability. Hence Step S4.4 is completed. □

Altogether, Steps S4.1– S4.4 conclude the proof of (ii) in Theorem 5.

Next, we give a sketch of the proof of (i), which describes the asymptotic power performance of the test under the global alternative with fixed $C_n \equiv C$. Let

$$\tilde{\theta} = (\tilde{\beta}; \tilde{\gamma}) = \arg\min_{\theta} E\left\{Y - \bar{W}E^{-1}[\bar{W}\bar{W}^{\mathrm{T}}]E[\bar{W}g(\beta^{\mathrm{T}}X, \gamma)]\right\}^2$$

which is different from the true parameter θ_0. Here $\bar{W}$ is a vector consisting of polynomials of W. In this case, $Z = B^{\mathrm{T}}W$ and $\tilde{b} = \tilde{\beta}/\|\tilde{\beta}\|$. Then, for fixed $C_n \equiv C$,

$$E[Y - r(\tilde{b}^{\mathrm{T}}W, \tilde{\theta})|Z] = E[CG(B^{\mathrm{T}}X) + r(b^{\mathrm{T}}W, \theta_0) - r(\tilde{b}^{\mathrm{T}}W, \tilde{\theta})|Z] := \tilde{\Delta}(Z).$$

We can obtain that V_n tends to a positive constant $E[\tilde{\Delta}^2(Z)f_Z(Z)]$ in probability. Similarly, we can also prove that $\hat{\tau}$ converges to a positive constant.

We then have that $V_n/\hat{\tau}$ converges in probability to a positive constant. That is, the test statistic $nh^{1/2}V_n$ goes to infinity at the rate of order $nh^{1/2}$. The proof is finished. □

S5 Proof of Theorem 1

As the arguments used for proving Theorem 5 with $C_n = 0$, the results $\|\hat{B} - B\| = O_p(1/\sqrt{n})$ and $\hat{\beta}_0 - \beta = O_p(1/\sqrt{n})$ are applicable for proving this theorem, we then omit most of the details, but focus on the bias term. The terms $\bar{f}_{N(j)}$, $Q_{k(j)}$, $k = 1, 2, 3$ and $j = 1, 2$ in the proof of Theorem 5 are replaced by

$$\bar{f}_N(x, b) = \frac{1}{N}\sum_{s=1}^{N} M_v(x - b^{\mathrm{T}}\tilde{w}_s), \tag{S5.1}$$

and

$$\begin{aligned} Q_1(x, b) &= \frac{1}{N}\sum_{s=1}^{N} M_v(x - b^{\mathrm{T}}\tilde{w}_s)(r_i - \tilde{r}_s), \\ Q_2(x, b) &= \frac{1}{N}\sum_{s=1}^{N} M_v(x - b^{\mathrm{T}}\tilde{w}_s)(\tilde{r}_s - \tilde{g}_s), \\ Q_3(x, b) &= \frac{1}{N}\sum_{s=1}^{N} M_v(x - b^{\mathrm{T}}\tilde{w}_s)(\tilde{g}_s - \hat{\tilde{g}}_s). \end{aligned} \tag{S5.2}$$

Using the same decomposition as in the proof of Step S4.4, we also have a term similar to I_{10} with the conditional expectation as

$$I_{10} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i - z_j)Q_2(z_i, b_0)Q_2(z_j, b_0)\frac{1}{\bar{f}_N(z_i, b_0)\bar{f}_N(z_j, b_0)}$$

and

$$E[I_{10}|\tilde{\eta}_s, \tilde{z}_s, \tilde{\eta}_t, \tilde{z}_t] \;=\; \frac{1}{N^2}\sum_{s=1}^{N}\sum_{t=1}^{N}\frac{1}{h}K(\frac{\tilde{z}_s-\tilde{z}_t}{h})\tilde{\eta}_s\tilde{\eta}_t(1+o_p(1)).$$

Separate the summands with $s \neq t$ and $s = t$ to write the leading term in the above expression as the sum of the following two terms.

$$I_{101}^{*} = \frac{1}{N^2}\sum_{s=1}^{N}\sum_{t\neq s}^{N}\frac{1}{h}K(\frac{\tilde{z}_s-\tilde{z}_t}{h})\tilde{\eta}_s\tilde{\eta}_t, \quad I_{102}^{*} = \frac{1}{N^2}\sum_{s=1}^{N}\frac{1}{h}K(0)\tilde{\eta}_s^2.$$

Since K is symmetric, I_{101}^{*} can be written as an U-statistic with the kernel

$$H_n((\tilde{z}_s, \tilde{\eta}_s), (\tilde{z}_t, \tilde{\eta}_t)) = \frac{1}{h}K(\frac{\tilde{z}_s-\tilde{z}_t}{h})\tilde{\eta}_s\tilde{\eta}_t.$$

Further,

$$E[H_n((\tilde{z}_s, \tilde{\eta}_s), (\tilde{z}_t, \tilde{\eta}_t))|(\tilde{z}_s, \tilde{\eta}_s)] = \frac{1}{h}\tilde{\eta}_s E\{K(\frac{\tilde{z}_s-\tilde{z}_t}{h}) \times E[\tilde{\eta}_t|\tilde{z}_t]\} = 0.$$

Thus the U-statistic I_{101}^{*} is degenerate. By Central Limit Theorem for degenerate U-statistic (see, Hall (1984)),

$$Nh^{1/2}I_{101}^{*} \rightarrow_D N(0, 2\int K^2(u)du\int(\xi^2(z))^2 f_Z^2(z)dz).$$

Hence $nh^{1/2}I_{101}^{*} \rightarrow_D N(0, \lambda^{-2}\tau_3)$, where τ_3 is defined in (S4.9). Further, the fact that $NhEI_{102}^{*} = K(0)E[\xi^2(Z)]$ implies that $nh^{1/2}EI_{102}^{*} \rightarrow \infty$, which results in the asymptotic bias in $\tilde{V}_n$. $\square$

S6 Proof of Theorem 3

When $N/n \to 0$, $\hat{\theta}_0$ and $\hat{B}(\hat{q})$ are $\sqrt{N}$ consistent estimates of θ_0 and B, respectively. Again as the decompositions used in the proof of Theorem 5 are applicable for proving this theorem, we give only a sketch of the proof of (i) here. Put $C_n = 0$ in the proof of Theorem 5. We only consider I_1, $V_{n2,1}$, and I_{10}. As $(Nv^{1/2})/(nh^{1/2}) \to 0$, $Nv^{1/2}I_{1,1}$ in Step S4.1 is $o_p(1)$. In addition, $Nh^2 \to \infty$ leads to $Nv^{1/2}I_{1,2} = o_p(1)$. Thus $Nv^{1/2}I_1 = o_p(1)$. For $V_{n2,1}$, following the proof of Step S4.2, we obtain that $Nv^{1/2}I_4 = o_p(1)$, $Nv^{1/2}I_5 = o_p(1)$, $Nv^{1/2}I_6 = o_p(1)$. These imply that $Nv^{1/2}V_{n2} = o_p(1)$. Recalling the notation in (S4.1), (S4.2), (S5.1) and (S5.2), I_{10} can be written as

$$I_{10} = \frac{1}{n(n-1)N^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_h(z_i - z_j)Q_2(z_i, b_0)Q_2(z_j, b_0)\frac{1}{\bar{f}_N(z_i, b_0)\bar{f}_N(z_j, b_0)}.$$

Again define its conditional expectation as

$$\begin{aligned} I_{10}^* &= E[I_{10}|\tilde{z}_s, \tilde{\eta}_s, \tilde{z}_t, \tilde{\eta}_t] \\ &= \frac{1}{N^2}\sum_{s=1}^{N}\sum_{t=1}^{N}\tilde{\eta}_s\tilde{\eta}_t \int\int \frac{1}{h}K(\frac{z_i - z_j}{h})\frac{1}{v}M(\frac{z_i - \tilde{z}_s}{v})\frac{1}{v}M(\frac{z_j - \tilde{z}_t}{v})dz_i dz_j. \end{aligned}$$

Thus,

$$\begin{aligned} &\int\int \frac{1}{h}K(\frac{z_i - z_j}{h})\frac{1}{v}M(\frac{z_i - \tilde{z}_s}{v})\frac{1}{v}M(\frac{z_j - z_t}{v})dz_i dz_j \\ =& \int\int \frac{1}{h}K(u)\frac{1}{v}M(\frac{hu + z_j - \tilde{z}_s}{v})\frac{1}{v}M(\frac{z_j - \tilde{z}_t}{v})d(z_j + uh)dz_j \end{aligned}$$

$$= \int \frac{1}{v}M(\frac{z_j - \tilde{z}_s}{v})\frac{1}{v}M(\frac{z_j - \tilde{z}_t}{v})dz_j + \int \frac{1}{v}M''(\frac{z_j - \tilde{z}_s}{v})\frac{h^2}{v^2}\frac{1}{v}M(\frac{z_j - \tilde{z}_t}{v})dz_j.$$

Then we have $I_{10}^* = (I_{101} + I_{102})(1 + o_p(1))$ where

$$I_{101} = \frac{1}{N^2}\sum_{s=1}^{N}\sum_{t\neq s}^{N}\tilde{\eta}_s\tilde{\eta}_t \int \frac{1}{v}M(\frac{z_j - \tilde{z}_s}{v})\frac{1}{v}M(\frac{z_j - \tilde{z}_t}{v})dz_j,$$
$$I_{102} = \frac{1}{N^2}\sum_{s=1}^{N}\tilde{\eta}_s^2 \int \frac{1}{v}M(\frac{z_j - \tilde{z}_s}{v})\frac{1}{v}M(\frac{z_j - \tilde{z}_s}{v})dz_j.$$

Rewrite I_{101} as

$$2\sum_{s=2}^{N}\sum_{t<s}^{N}\tilde{\eta}_s\tilde{\eta}_t\frac{1}{N^2}\int \frac{1}{v}M(\frac{z_j - \tilde{z}_s}{v})\frac{1}{v}M(\frac{z_j - \tilde{z}_t}{v})dz_j.$$

By Theorem 1 of Hall (1984), $Nv^{1/2}I_{101} \to_D N(0, \tilde{\tau})$, where

$$\tilde{\tau} = 2\int \Big(\int M(u)M(u+v)du)^2 dv \int (\xi^2(z)\Big)^2 f_Z^2(z)dz,$$

We also have in probability

$$NvI_{102} \to_p E[\int \frac{1}{v}M(\frac{z_j - \tilde{z}_s}{v})M(\frac{z_j - \tilde{z}_s}{v})dz_j\tilde{\eta}_s^2] = \int M^2(u)duE[\xi^2(z)].$$

ThenWe have $Nv^{1/2}\{I_{10}^* - \nu\} \to_D N(0, \tilde{\tau})$. We can further prove that

$$E[(I_{10} - I_{10}^*)^2] = O_p(\frac{1}{N^2nv}) = o_p(\frac{1}{N^2v}).$$

Hence $Nv^{1/2}\{I_{10} - \nu\} \to_D N(0, \tilde{\tau})$. This completes the proof of Theorem 3.

□

Bibliography

Guo, X., Wang, T. and Zhu, L.X. (2016). Model checking for parametric single-index models: a dimension reduction model-adaptive approach. *J. R. Stat. Soc. Series B Stat. Methodol. 78*, pp.1013–1035.

Hall, P. (1984). Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators. *J. Multivariate Anal. 14*, pp.1–16. **14**, 1–16.

Lee, L.F. and Sepanski, J.H. (1995). Estimation of Linear and Nonlinear Errors-in-Variables Models using Validation Data. *J. Am. Statist. Assoc. 90*, pp. 130–140.

Li, B., Wen, S. and Zhu, L.X. (2008). On a projective resampling method for dimension reduction with multivariate responses. *J. Am. Statist. Assoc. 103*, pp. 1177–1186.

Stute, W., Xue, L.G. and Zhu, L.X. (2007). Empirical Likelihood Inference in Nonlinear Errors-in-Covariables Models with Validation Data. *J. Am. Statist. Assoc. 102*, pp. 332–346.

Zheng, J.X. (1996). A Consistent Test of Functional Form via Nonparametric Estimation Technique. *J. Econometrics. 75*, pp. 263–289.

Zhu, L.P., Wang, T., Zhu, L.X. and Ferr, L. (2010a). Sufficient Dimension Reduction through Discretization-Expectation Estimation. *Biometrika. 97*, pp. 295–304.

Table 1: Empirical sizes and powers of $\tilde{T}_n$ of H_0 vs. H_{1k}, $k = 1, 2, 3$ in **Study 1**.

	a	p=2		p=8		p=2		p=8	
$\lambda = 4$		$\Sigma = \Sigma_1$		$\Sigma = \Sigma_1$		$\Sigma = \Sigma_2$		$\Sigma = \Sigma_2$	
H11		n=100	n=200	n=100	n=200	n=100	n=200	n=100	n=200
$\tilde{T}_n$	0	0.0485	0.0520	0.0440	0.0525	0.0440	0.0510	0.0485	0.0460
	0.1	0.0645	0.0760	0.0505	0.0865	0.0790	0.1300	0.1070	0.1615
	0.2	0.1130	0.2335	0.1230	0.2210	0.2010	0.4135	0.2720	0.6240
	0.3	0.2530	0.5205	0.2245	0.4975	0.4110	0.7900	0.5845	0.9500
	0.4	0.4365	0.8055	0.3800	0.7980	0.6945	0.9720	0.8125	0.9930
	0.5	0.6475	0.9495	0.5715	0.9360	0.8545	0.9995	0.9280	1.0000
H12		n=100	n=200	n=100	n=200	n=100	n=200	n=100	n=200
$\tilde{T}_n$	0	0.0445	0.0490	0.0500	0.0515	0.0555	0.0480	0.0475	0.0410
	0.1	0.0705	0.0825	0.0625	0.0790	0.0635	0.0855	0.0695	0.0820
	0.2	0.1375	0.2280	0.1130	0.2245	0.1425	0.2235	0.1055	0.1880
	0.3	0.2805	0.4830	0.2280	0.4630	0.2545	0.4335	0.1995	0.3615
	0.4	0.4415	0.7750	0.3700	0.7410	0.4165	0.7050	0.3120	0.6335
	0.5	0.6315	0.9250	0.5875	0.9165	0.5705	0.8935	0.4650	0.8275
H13		n=100	n=200	n=100	n=200	n=100	n=200	n=100	n=200
$\tilde{T}_n$	0	0.0455	0.0530	0.0585	0.0455	0.0475	0.0565	0.0500	0.0485
	0.1	0.0605	0.0910	0.0665	0.0805	0.0765	0.0965	0.0590	0.0725
	0.2	0.1360	0.2420	0.1100	0.2240	0.1100	0.1980	0.0880	0.1570
	0.3	0.2680	0.4595	0.2090	0.4440	0.2120	0.4065	0.1335	0.2905
	0.4	0.3750	0.6920	0.3365	0.6405	0.3375	0.6135	0.1910	0.4665
	0.5	0.5520	0.8730	0.4400	0.8375	0.4605	0.7775	0.2685	0.5910