MODEL CHECKING IN LARGE-SCALE DATA SET VIA STRUCTURE-ADAPTIVE-SAMPLING

Yixin Han[1], Ping Ma[2], Haojie Ren[3], and Zhaojun Wang[1]

[1] *School of Statistics and Data Science, LPMC & KLMDASR, Nankai University, Tianjin, P.R. China*
[2] *Department of Statistics, University of Georgia, Athens, GA, USA*
[3] *School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, P.R. China*

Supplementary Material

This Supplementary Material contains the proofs of several technical lemmas, the relevant proof of estimated dimension reduction direction, and some additional simulation results.

S1. Useful lemmas

The first lemma is a standard Bernstein's inequality.

Lemma S.1 (Bernstein's inequality)**.** *Let $Y_1, \ldots, Y_n$ be independent centered random variables a.s. bounded by $A < \infty$ in absolute value. Let $\sigma^2 = n^{-1}\sum_{i=1}^n \mathbb{E}(Y_i^2)$. Then for all $t > 0$,*

$$\Pr\left(\sum_{i=1}^n Y_i > t\right) \le \exp\left(-\frac{t^2}{2n\sigma^2 + 2At/3}\right).$$

The next lemma is a well-known projection result for U-statistic.

Lemma S.2 (Projection of U-statistic)**.** *Let $z_1, \ldots, z_n$ be an independent and identically random variable, $H_n(z_1, z_2)$ be an order two kernel of the U-statistics $U_n = \{n(n-1)\}^{-1}\sum_{i\neq j} H_n(z_i, z_j)$. Let $r_n(z_i) = \mathbb{E}\{H_n(z_i, z_j)|z_i\}$ be the projection on z_i. If* *we provide* $\mathbb{E}\left\{H_n^2(z_i,z_j)\right\}=o(n)$, *then we have*

$$U_n=\mathbb{E}\left\{r_n(z_i)\right\}+o_p(1).$$

Lemma S.3 is a direct result for Nadaraya-Watson estimator in Ren et al. (2020).

Lemma S.3 (Nadaraya-Watson estimator)**.** *Suppose the condition in Corollary 1 all hold. Under the "singular" local alternative* (2.6)*, the Nadaraya-Watson estimator* $\widehat{M}(\omega)$ *of* $M(\omega)$ *satisfies*

$$\sup_{\omega\in\Omega_n}\left|\widehat{M}(\omega)-M(\omega)\right|=O_p\left(h_f^2\delta_n'+\sqrt{\frac{a_n\log n_0}{n_0h_f}}\right).$$

S2. Proofs of Lemmas

Proof of Lemma A.1.

$W_n(\boldsymbol{\varepsilon},\boldsymbol{\varepsilon})$ can be regarded as a U-statistic with a kernel

$$H_n(z_i,z_j)=\frac{1}{h}K\left(\frac{\omega_i-\omega_j}{h}\right)\frac{\varepsilon_i}{\sqrt{f(\omega_i)}}\frac{\varepsilon_j}{\sqrt{f(\omega_j)}},$$

where $z_i=\{\omega_i,\varepsilon_i\}$. Under $\mathbb{H}_0$, $\mathbb{E}(\varepsilon_i\mid\omega_i)=0$. Thus, we can verify that

$$\mathbb{E}\left\{H_n(z_i,z_j)\mid z_i\right\}=\frac{\varepsilon_i}{h\sqrt{f(\omega_i)}}\mathbb{E}\left\{\frac{1}{\sqrt{f(\omega_j)}}K\left(\frac{\omega_i-\omega_j}{h}\right)\mathbb{E}(\varepsilon_j\mid\omega_j)\right\}=0,$$

this implies that $W(\boldsymbol{\varepsilon},\boldsymbol{\varepsilon})$ is a degenerate statistic of order two. By a similar proof of Lemma 3.3 in Zheng (1996) with the technique provided in Hall (1984), it is easy to obtain $nh^{1/2}W_n(\boldsymbol{\varepsilon},\boldsymbol{\varepsilon})/\sigma_V\xrightarrow{\mathcal{L}}\mathcal{N}(0,1)$, where $\sigma_V^2=2\sigma^4|\Omega|\int K^2(u)\mathrm{d}u$. □

Proof of Lemma A.2.

By Assumption 5 and the Bernstein's inequality (Lemma S.1), we can show that

$$|\hat{g}_k(\nu) - g_k^*(\nu)| \leq \sup_{\nu\in\Gamma}\left|R_k(\nu)b^2 + n^{-1}H_k(\nu)\sum_{i=1}^{n}\phi_k(\mathbf{X}_i)Q_b\left(u_{ki}-\nu\right)\varepsilon_i\right| + O_p(n^{-1/2})$$

$$= O_p\left\{b^2 + (nb/\log n)^{-1/2} + n^{-1/2}\right\}.$$

By Assumptions 4, it suffices to show that

$$\begin{aligned} G(\mathbf{X};\widehat{\boldsymbol{\beta}},\widehat{\mathbf{g}}) - G(\mathbf{X};\boldsymbol{\beta}^*,\mathbf{g}^*) &= \nabla G_{\boldsymbol{\beta}}^\top(\mathbf{X};\widetilde{\boldsymbol{\beta}},\widetilde{\mathbf{g}})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*) + \nabla G_{\mathbf{g}}^\top(\mathbf{X};\widetilde{\boldsymbol{\beta}},\widetilde{\mathbf{g}})(\widehat{\mathbf{g}}-\mathbf{g}^*) \\ &= O_p(n^{-1/2}) + O_p\left\{b^2 + (nb/\log n)^{-1/2} + n^{-1/2}\right\} \\ &= O_p\left\{b^2 + (nb/\log n)^{-1/2} + n^{-1/2}\right\}, \end{aligned}$$

here $\widetilde{g}_k(\nu)$ lies between $g_k(\nu)$ and $\widehat{g}_k(\nu)$, $k = 1, \ldots, q$, and $\widetilde{\boldsymbol{\beta}}$ lies between $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}$. By Assumption 4, the first derivatives of $G(\cdot)$ with respect to $\boldsymbol{\beta}$ and $\mathbf{g}$ are bounded. Then, the result is proved from the above discussion. □

Proof of Lemma A.3.

Note that for any fixed $\boldsymbol{\theta}$, $\mathbb{E}\{W_n(\boldsymbol{\varepsilon},\boldsymbol{\Upsilon}^*)\} = 0$ because $\mathbb{E}(\varepsilon_i \mid \omega_i) = 0$ under $\mathbb{H}_0$. Then, we calculate its second-order moment

$$\begin{aligned} \mathbb{E}\left\{W_n^2(\boldsymbol{\varepsilon},\boldsymbol{\Upsilon}^*)\right\} = \mathbb{E}\Bigg\{&\frac{1}{n^2(n-1)^2h^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{i'=1}^{n}\sum_{j'\neq i'}^{n}\frac{\varepsilon_i}{\sqrt{f(\omega_i)}}\frac{\Upsilon_j^*}{\sqrt{f(\omega_j)}} \\ &\cdot\frac{\varepsilon_{i'}}{\sqrt{f(\omega_{i'})}}\frac{\Upsilon_{j'}^*}{\sqrt{f(\omega_{j'})}}K\left(\frac{\omega_i-\omega_j}{h}\right)K\left(\frac{\omega_{i'}-\omega_{j'}}{h}\right)\Bigg\}. \end{aligned}$$

Since $\mathbb{E}(\varepsilon_i\varepsilon_{i'}) \neq 0$ if and only if $i = i'$, we have

$$\begin{aligned} \mathbb{E}\left\{W_n^2(\boldsymbol{\varepsilon},\boldsymbol{\Upsilon}^*)\right\} &= \frac{n(n-1)^2\sigma^2}{n^2(n-1)^2h^2}\mathbb{E}\left\{\frac{\Upsilon_j^*\Upsilon_{j'}^*}{f(\omega_i)\sqrt{f(\omega_j)}\sqrt{f(\omega_{j'})}}K\left(\frac{\omega_i-\omega_j}{h}\right)K\left(\frac{\omega_i-\omega_{j'}}{h}\right)\right\} \\ &\leq \frac{\sigma^2}{nh^2}\mathbb{E}\left\{\frac{1}{f(\omega_i)\sqrt{f(\omega_j)}\sqrt{f(\omega_{j'})}}K\left(\frac{\omega_i-\omega_j}{h}\right)K\left(\frac{\omega_i-\omega_{j'}}{h}\right)\right\}\left\{\sup_{\nu\in\Gamma}\left|\Upsilon_j^*\right|\right\}^2 \\ &= O(n^{-1})\cdot O\left[\left\{b^2 + (nb/\log n)^{-1/2} + n^{-1/2}\right\}^2\right], \end{aligned}$$

where the last inequality is due to Lemma A.2. Based on the bandwidth condition in Assumption 6 and Chebyshev's inequality, we have $W_n(\boldsymbol{\varepsilon}, \boldsymbol{\Upsilon}^*) = o_p(n^{-1}h^{-1/2})$.

By Assumption 4, we have the expansion $\Upsilon_i^* = \nabla G_{\widetilde{\boldsymbol{\beta}}}^\top(\mathbf{X}_i; \widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{g}})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \nabla G_{\widetilde{\mathbf{g}}}^\top(\mathbf{X}_i; \widetilde{\boldsymbol{\beta}}, \widetilde{\mathbf{g}})(\widehat{\mathbf{g}} - \mathbf{g}_0)$. Substituting Υ_i^* into $W_n(\boldsymbol{\Upsilon}^*, \boldsymbol{\Upsilon}^*)$ and re-expresses it as

$$
\begin{aligned}
W_n(\boldsymbol{\Upsilon}^*, \boldsymbol{\Upsilon}^*) = &\frac{1}{n(n-1)h} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\nabla G_{\widetilde{\mathbf{g}}}^\top \left\{\widehat{\mathbf{g}}(\mathbf{X}_i) - \mathbf{g}_0(\mathbf{X}_i)\right\}}{\sqrt{f(\omega_i)}} \frac{\nabla G_{\widetilde{\mathbf{g}}}^\top \left\{\widehat{\mathbf{g}}(\mathbf{X}_j) - \mathbf{g}_0(\mathbf{X}_i)\right\}}{\sqrt{f(\omega_j)}} K\left(\frac{\omega_i - \omega_j}{h}\right) \\
&+ \frac{2}{n(n-1)h} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\nabla G_{\widetilde{\mathbf{g}}}^\top \left\{\widehat{\mathbf{g}}(\mathbf{X}_i) - \mathbf{g}_{0i}(\mathbf{X}_i)\right\}}{\sqrt{f(\omega_i)}} \frac{\nabla G_{\widetilde{\boldsymbol{\beta}}}^\top \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)}{\sqrt{f(\omega_j)}} K\left(\frac{\omega_i - \omega_j}{h}\right) \\
&+ \frac{1}{n(n-1)h} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\nabla G_{\widetilde{\boldsymbol{\beta}}}^\top \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)}{\sqrt{f(\omega_i)}} \frac{\nabla G_{\widetilde{\boldsymbol{\beta}}}^\top \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)}{\sqrt{f(\omega_j)}} K\left(\frac{\omega_i - \omega_j}{h}\right) \\
:= &W_{n11} + 2W_{n12} + W_{n13}.
\end{aligned}
$$

For W_{n12}, we know that

$$
W_{n12} \leq \left\{ \frac{1}{n(n-1)h} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\nabla G_{\widetilde{\boldsymbol{\beta}}}^\top \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)}{\sqrt{f(\omega_i)}\sqrt{f(\omega_j)}} K\left(\frac{\omega_i - \omega_j}{h}\right) \nabla G_{\widetilde{\mathbf{g}}}^\top \sup_{\nu\in\Gamma} \left|\widehat{\mathbf{g}}(\mathbf{X}_i) - \mathbf{g}_0(\mathbf{X}_i)\right| \right\},
$$

the inequality holds since the kernel $K(\cdot)$, the derivatives $\nabla G_{\widetilde{\mathbf{g}}}$ and $G_{\widetilde{\boldsymbol{\beta}}}$, and density $f(\cdot)$ are positive and bounded functions. By Lemma A.2, we get $\sup_{\nu\in\Gamma} \left|\widehat{\mathbf{g}}(\mathbf{X}_i) - \mathbf{g}_0(\mathbf{X}_i)\right| = O_p\left\{b^2 + (nb/\log n)^{-1/2}\right\}$. Based on the bandwidth conditions in Assumption 6

$$
W_{n12} = O_p\left(n^{-1/2}\right) \cdot O_p\left\{b^2 + (nb/\log n)^{-1/2}\right\} = o_p\left(n^{-1}h^{-1/2}\right).
$$

For W_{n13}, it is easy to check that $W_{n13} = O_p\left(n^{-1}\right) = o_p\left(n^{-1}h^{-1/2}\right)$.

Next, we discuss the term W_{n11}. By the representation of link function $\mathbf{g} = (g_1, \ldots, g_q)^\top$, there exists a bounded function $A(\cdot)$ such that

$$
\widehat{g}_k(\nu) = \frac{n^{-1} \sum_{i=1}^n A(\mathbf{X}_i) Q_b\left(u_{ki} - \nu\right) Y_i}{n^{-1} \sum_{j=1}^n A(\boldsymbol{X}_j) Q_b\left(u_{kj} - \nu\right)} + O_p(n^{-1}),\ k = 1, \ldots, q. \tag{S.1}
$$

Let $C(\nu) = n^{-1} \sum\limits_{i=1}^{n} A(\mathbf{X}_i) Q_b\left(u_{ki} - \nu\right)$. Substituting (S.1) into W_{n11}

$$
\begin{aligned}
W_{n11} =& \frac{1}{n^2(n-1)^2 hb^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{s=1}^{n} \sum_{t \neq s}^{n} \left\{ \frac{1}{\sqrt{f(\omega_i)}\sqrt{f(\omega_j)} C(u_i) C(u_j)} K\left(\frac{\omega_i - \omega_j}{h}\right) \right. \\
& \left. \cdot \nabla G_{\tilde{\mathbf{g}}}^{\top} A(\mathbf{X}_i) Q\left(\frac{u_s - u_i}{b}\right)\left(Y_s - \mathbf{g}_0(\mathbf{X}_i)\right) \nabla G_{\tilde{\mathbf{g}}}^{\top} A(\mathbf{X}_j) Q\left(\frac{u_t - u_j}{b}\right)\left(Y_t - \mathbf{g}_0(\mathbf{X}_j)\right) \right\} + O_p(n^{-1}).
\end{aligned}
$$

Then, we need to prove $W_{n11} = o_p\left(n^{-1}h^{-1/2}\right)$ in the following two cases.

Case 1. {The indices i, j, s, t are all different}. Denote the expectation result as S_1. By the Assumption 5, the Lemma B.1 in Fan and Li (1996), and Lemma 2 and Lemma 3 in Robinson (1988), we have

$$
\begin{aligned}
S_1 =& \frac{1}{hb^2\sqrt{f(\omega_i)}\sqrt{f(\omega_j)} C(u_i) C(u_j)} \mathbb{E}\left[\mathbb{E}_s \left\{ \nabla G_{\tilde{\mathbf{g}}}^{\top} A(\mathbf{X}_i) Q\left(\frac{u_s - u_i}{b}\right)\left(Y_s - \mathbf{g}_0(\mathbf{X}_i)\right) \right\} \right. \\
& \left. \cdot \mathbb{E}_t \left\{ \nabla G_{\tilde{\mathbf{g}}}^{\top} A(\mathbf{X}_i) Q\left(\frac{u_t - u_j}{b}\right)\left(Y_t - \mathbf{g}_0(\mathbf{X}_j)\right) \right\} \mid K\left(\frac{\omega_i - \omega_j}{h}\right) \right] \\
\leq & \frac{C_3 b^{2r}}{h} \mathbb{E}\left\{ D_{\mathbf{g}}(\boldsymbol{X}_i) D_{\mathbf{g}}(\mathbf{X}_j) K\left(\frac{\omega_i - \omega_j}{h}\right) \right\} \\
=& O(b^{2r}) = o(n^{-1}h^{-1/2}),
\end{aligned}
$$

where C_3 is a positive constant, and $D_{\mathbf{g}}(\cdot)$ is the rth bounded derivative of $\mathbf{g}(\cdot)$.

Case 2. {The indices i, j, s, t take no more than three different values}. Denote the expectation result as S_2. By the same conditions in Case 1, we have

$$
\begin{aligned}
S_2 =& \frac{1}{hb^2\sqrt{f(\omega_i)}\sqrt{f(\omega_j)} C(u_i) C(u_j)} \\
& \cdot \mathbb{E}\left[\mathbb{E}^2 \left\{ \nabla G_{\tilde{\mathbf{g}}}^{\top} A(\mathbf{X}_i) Q\left(\frac{u_i - u_j}{b}\right)\left(Y_i - \mathbf{g}_0(\mathbf{X}_j)\right) \right\} \mid K\left(\frac{\omega_i - \omega_j}{h}\right) \right] \\
\leq & \frac{C b^{2r}}{h} \mathbb{E}\left\{ D_{\mathbf{g}}(\mathbf{X}_i) D_{\mathbf{g}}(\mathbf{X}_j) K\left(\frac{\omega_i - \omega_j}{h}\right) \right\} \\
=& O(b^{2r}) = o(n^{-1}h^{-1/2}).
\end{aligned}
$$

Hence, $\mathbb{E}(W_{n11}) = S_1 + S_2 = o\left(n^{-1}h^{-1/2}\right)$. By similar discussion (we omit tedious process for simplicity), there is $\mathbb{E}(W_{n11}^2) = o\left(n^{-2}h^{-1}\right)$. Further, the application of Chebyshev's inequality yields that $W_{n11} = o_p\left(n^{-1}h^{-1/2}\right)$.

In summary, we arrive at the result that $W_n(\boldsymbol{\Upsilon}^*, \boldsymbol{\Upsilon}^*) = o_p(n^{-1}h^{-1/2})$. □

Proof of Lemma A.4.

Note that $W_n(\boldsymbol{\varepsilon}, \mathbf{L})$ be rewritten as a U-statistic with kernel

$$H_n(z_i, z_j) = \frac{1}{2h\sqrt{f(\omega_i)}\sqrt{f(\omega_j)}} K\left(\frac{\omega_i - \omega_j}{h}\right)\left\{\varepsilon_i L(\mathbf{X}_j) + \varepsilon_j L(\mathbf{X}_i)\right\}.$$

By the theory of non-degenerate U-statistic in Serfling (2009)

$$\begin{aligned}
&\mathbb{E}\left\{H_n^2(z_i, z_j)\right\} \\
&\leq 2\mathbb{E}\left\{\frac{1}{2h}\frac{\varepsilon_i L(\mathbf{X}_j)}{\sqrt{f(\omega_i)}\sqrt{f(\omega_j)}} K\left(\frac{\omega_i - \omega_j}{h}\right)\right\}^2 + 2\mathbb{E}\left\{\frac{1}{2h}\frac{\varepsilon_j L(\mathbf{X}_i)}{\sqrt{f(\omega_i)}\sqrt{f(\omega_j)}} K\left(\frac{\omega_j - \omega_i}{h}\right)\right\}^2 \\
&= \frac{1}{h^2}\int \frac{\sigma^2}{f(\omega_i)f(\omega_j)} K^2\left(\frac{\omega_i - \omega_j}{h}\right) L^2(\mathbf{X}_j) f(\omega_i) f(\omega_j) \mathrm{d}\omega_i \mathrm{d}\omega_j \\
&\leq \frac{\sigma^2}{h}\int \varphi^2(\mathbf{X}_j) K^2(u) \mathrm{d}\omega_i \mathrm{d}u \\
&= O(h^{-1}) = o(n),
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left\{H_n(z_i, z_j) \mid z_i\right\} &= \frac{\varepsilon_i}{2h\sqrt{f(\omega_i)}}\mathbb{E}\left[\mathbb{E}\left\{K\left(\frac{\omega_i - \omega_j}{h}\right)\frac{1}{\sqrt{f(\omega_j)}}\mathbb{E}\left\{L(\mathbf{X}_j)\right\} \mid \omega_i, \omega_j\right\}\right] \\
&= \frac{\varepsilon_i}{2h\sqrt{f(\omega_i)}}\mathbb{E}\left[K\left(\frac{\omega_i - \omega_j}{h}\right)\frac{1}{\sqrt{f(\omega_j)}}\mathbb{E}\left\{L(\mathbf{X}_j)\right\}\right] \\
&= \frac{\varepsilon_i}{2h\sqrt{f(\omega_i)}}\int \mathbb{E}\left\{L(\mathbf{X}_j \mid \omega_i - hu)\right\}\sqrt{f(\omega_i - hu)} K(u) h \mathrm{d}u \\
&= \frac{\varepsilon_i \mathbb{E}\left\{L(\mathbf{X}_j)\right\}}{2} + l_n(z_i).
\end{aligned}$$

With these results, we have the projection of statistic $W_n(\boldsymbol{\varepsilon}, \mathbf{L})$ (Lemma S.2) as

$$\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{E}\left\{H_n(z_i, z_j) \mid z_i\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\mathbb{E}\left\{L(\mathbf{X}_j)\right\} + \frac{2}{\sqrt{n}}\sum_{i=1}^{n} l_n(z_i) = O_p(1).$$

Note that $\widehat{f}(\omega - hu) = \widehat{f}(\omega) - \widehat{f}'(\omega)hu$ and $\int f'(\omega)^2(hu)^2K(u)\mathrm{d}u = O(h^2)$. Thus, $\mathbb{E}\left\{l_n^2(z_i)\right\} = O(h^2) \to 0$. As a result, we have $W_n(\boldsymbol{\varepsilon}, \mathbf{L}) = O_p(n^{-1/2})$. □

Proof of Lemma A.5.

Note that $f(\omega) \propto M^2(\omega)$. It is straightforward to verify this result under (2.3) by using Lemma S.3. □

S3. Proofs of the asymptotic results with an estimated $\boldsymbol{\theta}$

Lemma S.4. *Suppose the Assumptions 1–7 hold. Given $L(\cdot)$ is a continuously differentiable function, which satisfies $|L(\mathbf{X})| \le \varphi(\mathbf{X})$ for all $\mathbf{X} \in \mathbb{R}^p$ and $\mathbb{E}\left\{\varphi^2(\mathbf{X})\right\} < \infty$. If $h \to 0$ and $n_0h^{3/2} \to \infty$, then under the null hypothesis $\mathbb{H}_0$, the following result holds with $\widehat{\boldsymbol{\theta}}$*

$$W_n(\boldsymbol{\varepsilon}, \mathbf{L}, \widehat{\boldsymbol{\theta}}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\varepsilon_i}{\sqrt{f(\widehat{\omega}_i)}}K_h(\widehat{\omega}_i - \widehat{\omega}_j)\frac{L(\mathbf{X}_j)}{\sqrt{f(\widehat{\omega}_j)}} = O_p(n^{-1/2}h^{-1/4}).$$

Proof. This result is a variation of Lemma A.4 with an estimated dimension reduction direction. By Assumption 7, we know that $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_p(n_0^{-1/2})$. A new error term is involved by $\widehat{\boldsymbol{\theta}}$

$$\begin{aligned} W_K &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\varepsilon_i}{\sqrt{f(\omega_i)}\sqrt{f(\omega_j)}}\frac{1}{h}L(\mathbf{X}_j)\left\{K\left(\frac{\widehat{\omega}_i - \widehat{\omega}_j}{h}\right) - K\left(\frac{\omega_i - \omega_j}{h}\right)\right\} \\ &= \frac{1}{n(n-1)h^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\varepsilon_i}{\sqrt{f(\omega_i)}}\frac{L(\mathbf{X}_j)}{\sqrt{f(\omega_j)}}K'\left(\frac{\widetilde{\omega}_i - \widetilde{\omega}_j}{h}\right)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top(\mathbf{X}_i - \mathbf{X}_j), \end{aligned}$$

the second formula holds by mean value theorem. Here $\widetilde{\omega}_i = \widetilde{\boldsymbol{\theta}}^\top \mathbf{X}_i \in \Omega$ with $\widetilde{\boldsymbol{\theta}}_j \in \left[\min\{\widehat{\boldsymbol{\theta}}_j, \boldsymbol{\theta}_j\}, \max\{\widehat{\boldsymbol{\theta}}_j, \boldsymbol{\theta}_j\}\right]$, $i = 1, \dots n; j = 1, \dots p$. By Assumption 3, we know the derivative of $K(\cdot)$ with respect to $\boldsymbol{\theta}$ is bounded, then we assert that replacing $\widetilde{\boldsymbol{\theta}}$ by $\boldsymbol{\theta}$ does not impact the convergence rate of W_K. That is

$$\frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{\varepsilon_i}{\sqrt{f(\omega_i)}} \frac{L(\mathbf{X}_j)}{\sqrt{f(\omega_j)}} K\left(\frac{\widetilde{\omega}_i - \widetilde{\omega}_j}{h}\right)(\mathbf{X}_i - \mathbf{X}_j),$$

can be rewritten as a U-statistic. Then, we can similarly show that this term is of order $O_p(n^{-1/2})$. Under the condition $n_0 h^{3/2} \to \infty$, the convergence rate of $W_n(\boldsymbol{\varepsilon}, \mathbf{L}, \widehat{\boldsymbol{\theta}})$ is $O_p(n_0^{-1/2} h^{-1} n^{-1/2}) = O_p(n^{-1/2} h^{-1/4} n_0^{-1/2} h^{-3/4}) = o_p(n^{-1/2} h^{-1/4})$. □

Proof of Theorem 3.

Theorem 3 is a direct extension of the results in Lemma A.1 and Proposition A.1 as long as the difference incurred by the estimated direction in kernel function and sampling density can be controlled. For the difference in kernel function $K(\cdot)$, we can refer to the derivation in Lemma S.4. Next, we focus on the differences that appear in the density $f(\cdot)$. Take Lemma A.1 as an example, the error term is involved in $f(\cdot)$

$$W_f = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{\varepsilon_i}{\sqrt{f(\omega_i)}} \frac{\varepsilon_j}{\sqrt{f(\omega_j)}} K_h(\widehat{\omega}_i - \widehat{\omega}_j) \left\{ \frac{\sqrt{f(\omega_i)}\sqrt{f(\omega_j)}}{\sqrt{\widehat{f}(\widehat{\omega}_i)}\sqrt{\widehat{f}(\widehat{\omega}_j)}} - 1 \right\}.$$

By the uniform convergence rate of kernel density estimator in Silverman (1978) and Lemma A.5, $\sup_\omega |\widehat{f}(\omega) - f(\omega)| = O_p\left\{h_f^2 + (n_0 h_f / \log n_0)^{-1/2}\right\}$. By Assumption 7, the standard derivations yield that

$$\left[\sup_{\omega_i, \omega_j} \left\{ \frac{\sqrt{f(\omega_i)}\sqrt{f(\omega_j)}}{\sqrt{\widehat{f}(\widehat{\omega}_i)}\sqrt{\widehat{f}(\widehat{\omega}_j)}} - 1 \right\}^2\right]^{1/2} = O_p\left\{n_0^{-1/2} + h_f^2 + (n_0 h_f / \log n_0)^{-1/2}\right\}. \quad \text{(S.2)}$$

Note that $\mathbb{E}(W_f)=0$. We next consider its second-order moment

$$\mathbb{E}(W_f^2)=\mathbb{E}\left\{\frac{1}{n^2(n-1)^2h^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{i'=1}^{n}\sum_{j'\neq i'}^{n}\frac{\varepsilon_i}{\sqrt{f(\omega_i)}}\frac{\varepsilon_j}{\sqrt{f(\omega_j)}}\frac{\varepsilon_{i'}}{\sqrt{f(\omega_{i'})}}\frac{\varepsilon_{j'}}{\sqrt{f(\omega_{j'})}}K\left(\frac{\widehat{\omega}_i-\widehat{\omega}_j}{h}\right)\right.$$
$$\left.\cdot K\left(\frac{\widehat{\omega}_{i'}-\widehat{\omega}_{j'}}{h}\right)\left(\frac{\sqrt{f(\omega_i)}\sqrt{f(\omega_j)}}{\sqrt{\widehat{f}(\widehat{\omega}_i)}\sqrt{\widehat{f}(\widehat{\omega}_j)}}-1\right)\left(\frac{\sqrt{f(\omega_{i'})}\sqrt{f(\omega_{j'})}}{\sqrt{\widehat{f}(\widehat{\omega}_{i'})}\sqrt{\widehat{f}(\widehat{\omega}_{j'})}}-1\right)\right\}.$$

Observe that $\mathbb{E}(\varepsilon_i\varepsilon_j\varepsilon_{i'}\varepsilon_{j'})\neq 0$ if and only if $i=i', j=j'$ or $i=j', j=i'$. Then, we can take a supremum for the last two terms in the above summation. By the result in (S.2), we have $\mathbb{E}(W_f^2)=O_p\left\{\log n_0/(n^2hn_0h_f)\right\}$. The application of Chebyshev's inequality yields $W_f=o_p(n^{-1}h^{-1/2})$. When replacing $\boldsymbol{\theta}$ with $\widehat{\boldsymbol{\theta}}$, the same results can be obtain in Lemma A.3.

By the same techniques discussed above, we get

$$\widehat{\sigma}_V^2=\frac{2h}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\varepsilon_i^2\varepsilon_j^2K_h^2(\omega_i-\omega_j)}{f(\omega_i)f(\omega_j)}+o_p(1),$$

where the main term is a U-statistic of order two with kernel function

$$H_n(z_i,z_j)=\frac{1}{h}K^2\left(\frac{\omega_i-\omega_j}{h}\right)\frac{\varepsilon_i^2}{f(\omega_i)}\frac{\varepsilon_j^2}{f(\omega_j)},$$

$$\begin{aligned}\mathbb{E}\left\{H_n(z_i,z_j)\right\}&=\frac{1}{h}\int K^2\left(\frac{\omega_i-\omega_j}{h}\right)\frac{\sigma^2}{f(\omega_i)}\frac{\sigma^2}{f(\omega_j)}f(\omega_i)f(\omega_j)\mathrm{d}\omega_i\mathrm{d}\omega_j\\&=\frac{\sigma^4}{h}\int K^2(u)\mathrm{d}uh\mathrm{d}\omega+o(1)\\&=\sigma^4|\Omega|\int K^2(u)\mathrm{d}u+o(1)\\&=\sigma_V^2/2+o(1).\end{aligned}$$

And $\mathbb{E}\left\{H_n^2(z_i,z_j)\right\}=o(h^{-1})=o(n)$. By Lemma S.2, we have $\widehat{\sigma}_V^2=\sigma_V^2+o_p(1)$.

Consequently, we can naturally get the conclusions in Theorem 3. □

S4. Addition simulations

To explore the influence of the dimension reduction methods and kernel functions in our proposed algorithm, we conduct the following studies by 500 replications on Scenarios I−III listed in Section 4. The results are reported in Table S1 and Table S2.

Table S1: Empirical sizes and powers (%) of the SAS procedure for different SDR methods under Scenarios I−III.

		MAVE			**SAVE**			**DR**		
Scenario	δ / structure	0.00	0.25	0.50	0.00	0.25	0.50	0.00	0.25	0.50
I	IID	5.8	97.0	99.8	7.0	97.4	100.0	5.6	94.4	100.0
	COR	6.4	11.8	65.2	6.8	12.6	66.6	4.0	12.0	64.8
II	IID	4.4	17.4	75.4	5.2	17.2	82.8	5.4	13.6	70.6
	COR	6.2	5.6	16.8	5.8	9.0	18.4	4.6	7.0	16.2
III	IID	5.2	9.0	31.8	6.6	7.4	10.2	5.2	6.2	13.0
	COR	7.0	4.4	15.2	5.6	6.8	10.4	6.4	6.2	10.8

Table S1 reveals that the dimension reduction methods are not very sensitive to our test procedure under three scenarios. In Table S2, the influence of different kernel functions on empirical sizes and powers of our SAS procedure is not obvious. As a result, we choose the MAVE dimension reduction method and Epanechnikov kernel function in our numerical analysis.

Table S2: Empirical sizes and powers (%) of the SAS procedure for different kernel functions under Scenarios I−III.

Scenario	structure \ δ	Epanechnikov 0.00	Epanechnikov 0.25	Epanechnikov 0.50	Triangular 0.00	Triangular 0.25	Triangular 0.50	Triweight 0.00	Triweight 0.25	Triweight 0.50
I	IID	5.8	97.0	99.8	5.6	96.2	99.8	5.6	95.6	100.0
	COR	6.4	11.8	65.2	6.8	12.0	62.8	5.4	12.0	56.6
II	IID	4.4	17.4	75.4	5.6	18.0	75.6	6.8	15.8	71.8
	COR	6.2	5.6	16.8	4.8	10.0	17.0	5.2	8.0	14.0
III	IID	5.2	9.0	31.8	5.0	7.6	28.0	4.6	6.8	26.8
	COR	7.0	4.4	15.2	7.6	6.2	16.0	6.6	6.2	13.2

References

Fan, Y. and Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica: Journal of the Econometric Society*, pages 865–890.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis*, **14**(1):1–16.

Ren, H., Zou, C., Chen, N., and Li, R. (2020). Large-scale datastreams surveillance via pattern-oriented-sampling. *Journal of the American Statistical Association*, pages 1–15.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954.

Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*. John Wiley & Sons.

Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics*, pages 177–184.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, **75**(2):263–289.