Online Supplementary Material for "Semiparametric estimation and variable selection for sparse single index models in increasing dimension"

Chaohua Dong[a] and Yundong Tu[b*]

[a]School of Statistics and Mathematics and

Research Center of Quantitative Economics

Zhongnan University of Economics and Law, China

[b]Guanghua School of Management and

Center for Statistical Science,

Peking University, Beijing, 100871, China

January 9, 2024

This supplementary document contains three parts. Part D contains the proofs of lemmas collected in Appendix A, Part E provides the verification of the conditions in Lemmas C.1 and C.2, and Part F presents some additional simulation results.

D Proofs of Lemmas in Appendix A

Proof of Lemma A.1. See Lemma B.1 and its proof in Dong et al. (2015). □

Proof of Lemma A.2. (1) The convergence is well known in the literature since g is smooth. The convergence rates for both the weighted super norm and the L^2 can be calculated similarly as Lemma C.1 in the supplement of Dong et al. (2016). (2) By virtue of Assumption 3.2, $\sup_{\boldsymbol{\theta}\in\Theta} E|\gamma_k(\boldsymbol{\theta}'\mathbf{x}_1)|^2 \leq C\|\gamma_k(v)\|_{L^2}^2 = o(k^{-\nu})$. □

*Correspondence to: Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, 100871, China. Tel: +86 10 62760219. E-mail: yundong.tu@gsm.pku.edu.cn.

Proof of Lemma A.3. (1) The matrix $\frac{1}{n}Z'_{(S)}WZ_{(S)}$ has elements $\frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_{\mathbf{p}}(\mathbf{x}_t)\mathcal{H}_{\mathbf{q}}(\mathbf{x}_t)w(\widetilde{\mathbf{x}}_t)$, and it is standard that $\frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_{\mathbf{p}}(\mathbf{x}_t)\mathcal{H}_{\mathbf{q}}(\mathbf{x}_t)w(\widetilde{\mathbf{x}}_t) - E[\mathcal{H}_{\mathbf{p}}(\mathbf{x}_1)\mathcal{H}_{\mathbf{q}}(\mathbf{x}_1)w(\widetilde{\mathbf{x}}_1)] = O_P(1/\sqrt{n})$ uniformly in $\mathbf{p}$ and $\mathbf{q}$, due to the uniform boundedness of $\mathcal{H}_{\mathbf{p}}(\cdot)\mathcal{H}_{\mathbf{q}}(\cdot)w(\cdot)$ over all $\mathbf{p}$ and $\mathbf{q}$ as well as the argument involved. See, e.g. Chapter One of Gautschi (2004, p. 27-31) or Szego (1975, p. 242). The assertion follows immediately.

(2) It holds similarly as (1).

(3) Note that $\frac{1}{n}Z'_{(S)}W(\mathbf{Y}-Z\boldsymbol{\beta}_0) = \frac{1}{n}Z'_{(S)}W\mathbf{e} + \frac{1}{n}Z'_{(S)}W\boldsymbol{\gamma}$. Here, the s-vector $\frac{1}{n}Z'_{(S)}W\mathbf{e}$ has elements $\frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_{\mathbf{p}}(\mathbf{x}_t)w(\widetilde{\mathbf{x}}_t)e_t$ and by Assumption 3.1 and uniform boundedness of $\mathcal{H}_{\mathbf{p}}(\mathbf{x}_t)w(\widetilde{\mathbf{x}}_t)$ over all $\mathbf{p}$ and t we have $\frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_{\mathbf{p}}(\mathbf{x}_t)w(\widetilde{\mathbf{x}}_t)e_t = O_P(n^{-1/2})$ uniformly. Thus, $\frac{1}{n}Z'_{(S)}W\mathbf{e} = O_P(\sqrt{s/n})$.

In addition, from Lemma A.2, $\|\frac{1}{n}Z'_{(S)}W\boldsymbol{\gamma}\|^2 \leq \frac{1}{n}\|\boldsymbol{\gamma}\|^2\|\frac{1}{n}Z_{(S)}W^2Z_{(S)}\| = O_P(k^{-\nu}s)$.

(4) Observe that after premultiplying the selection matrix, $QZ'_{(S)}W(\mathbf{Y}-Z\boldsymbol{\beta}_0) = QZ'_{(S)}W\mathbf{e}+QZ'_{(S)}W\boldsymbol{\gamma}$ becomes a vector of dimension s. The assertion follows immediately in view of the proof of (3). $\square$

E Verification of conditions in Lemmas C.1 and C.2

Verification of Conditions in Lemma C.1. Condition (1): Note that, ignoring unimportant constant, $F_{n(S_0)}(\boldsymbol{\beta}_{0(S_0)}) = \frac{1}{n}Z'_{(S_0)}W(\mathbf{Y}-Z_{(S_0)}\boldsymbol{\beta}_{0(S_0)})$. It has elements $\frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_{\mathbf{p}}(\mathbf{x}_t)w(\widetilde{\mathbf{x}}_t)[\gamma_k(\boldsymbol{\theta}_0'\boldsymbol{x}_t)+e_t]$ where $\mathbf{p}\mapsto j$ and $j\in S_0$. Due to the uniform boundedness of $\mathcal{H}_{\mathbf{p}}(\mathbf{x}_t)w(\widetilde{\mathbf{x}}_t)$ again over all $\mathbf{p}$ and the argument involved, and by virtue of Assumption 3.1 we can use the exponential-tail Bernstein inequality to have $\frac{1}{n}Z'_{(S_0)}W\mathbf{e} = O_P(\sqrt{s_0\log(K)/n})$ similar to Lemma C.2 in Fan and Liao (2014); and

$$\begin{aligned}\left(\frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_{\mathbf{p}}(\mathbf{x}_t)w(\widetilde{\mathbf{x}}_t)\gamma_k(\boldsymbol{\theta}_0'\mathbf{x}_t)\right)^2 \leq& \frac{1}{n^2}\sum_{t=1}^{n}[\mathcal{H}_{\mathbf{p}}(\mathbf{x}_t)w(\widetilde{\mathbf{x}}_t)]^2\sum_{t=1}^{n}[\gamma_k(\boldsymbol{\theta}_0'\mathbf{x}_t)]^2\\ \leq& C\frac{1}{n}\sum_{t=1}^{n}\gamma_k^2(\boldsymbol{\theta}_0'\mathbf{x}_t) = O_P(\|\gamma_k^2(u)\|^2)\\ =& O_P(k^{-\nu}),\end{aligned}$$

by Assumptions 3.2 and 3.3 as well as Lemma A.2, where we have used the fact that almost surely $\sum_{t=1}^{n}[\mathcal{H}_{\mathbf{p}}(\mathbf{x}_t)w(\widetilde{\mathbf{x}}_t)]^2 \leq Cn$ and $E[\gamma_k(\boldsymbol{\theta}_0'\mathbf{x}_t)]^2 \leq C\int\gamma_k^2(u)f_{\theta_0}(u)du \leq C\|\gamma_k(u)\|^2$. Hence, $\|F_{n(S_0)}(\boldsymbol{\beta}_{0(S_0)})\| = O_P(\sqrt{s_0\log(K)/n}+\sqrt{s_0}\|\gamma_k(u)\|)$.

With this order of $\|F_{n(S_0)}(\boldsymbol{\beta}_{0(S_0)})\|$, we then may conclude $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\| = O_P(\sqrt{s_0 \log(K)/n}+\sqrt{s_0}\|\gamma_k(u)\| + \sqrt{s_0}P_n'(\zeta_n))$ for the $\widehat{\boldsymbol{\beta}}$ in Lemmas C.1-C.2.

Condition (2) is mostly encountered in the literature. The readers are referred to Theorem 4.1 in Fan and Liao (2014) and Condition A.2 in Belloni et al. (2015).

Verification of Conditions in Lemma C.2. Let $\widehat{\boldsymbol{\beta}} \in \mathcal{V}$ be the minimizer of $Q_n(\boldsymbol{\beta}_{S_0})$. We shall show that there is a neighborhood of $\widehat{\boldsymbol{\beta}}$ in which for any $\boldsymbol{\beta} \notin \mathcal{V}$, the condition of (C.1) holds, that is, $L_n(\boldsymbol{\beta}_{S_0}) - L_n(\boldsymbol{\beta}) < \sum_{j\notin S_0} P_n(|\beta_j|)$. This is tantamount to showing $Q_n(\boldsymbol{\beta}_{S_0}) < Q_n(\boldsymbol{\beta})$.

Note that

$$\begin{aligned} L_n(\boldsymbol{\beta}_{S_0}) - L_n(\boldsymbol{\beta}) =& \frac{1}{n}[(\mathbf{Y} - Z\boldsymbol{\beta}_{S_0})'W(\mathbf{Y} - Z\boldsymbol{\beta}_{S_0}) - (\mathbf{Y} - Z\boldsymbol{\beta})'W(\mathbf{Y} - Z\boldsymbol{\beta})] \\ =& -2\frac{1}{n}(\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta})'Z'W(\mathbf{Y} - Z\boldsymbol{\beta}) + \frac{1}{n}(\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta})'Z'WZ(\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta}). \end{aligned}$$

In a small neighborhood of $\widehat{\boldsymbol{\beta}}$, $O(\widehat{\boldsymbol{\beta}}, r_n/K)$ say, where r_n is a sufficient small number determined later, $\sup_{\boldsymbol{\beta}\in O}\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\|_1 \leq r_n$ and

$$\begin{aligned} &\frac{1}{n}|(\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta})'Z'W(\mathbf{Y} - Z\boldsymbol{\beta})| + \frac{1}{n}(\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta})'Z'WZ(\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta}) \\ \leq& \frac{1}{n}|(\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta})'Z'WZ(\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta})| + \frac{1}{n}|(\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta})'Z'_{(S_0)}W(\mathbf{Y} - Z_{(S_0)}\boldsymbol{\beta}_{(S_0)})| \\ \leq& \|\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta}\|\|H_n(\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta})\| + \|\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta}\|_1\|F_{n(S_0)}(\boldsymbol{\beta}_1)\|_\infty + \|H_n\|_{\max}\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{S_0}\|_1\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_{S_0}\| \\ \leq& \|\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta}\|_1[Cr_n + r_n\|F_{n(S_0)}(\boldsymbol{\beta}_1)\| + C_1\|\boldsymbol{\beta}_1 - \widehat{\boldsymbol{\beta}}_{S_0}\|]. \end{aligned}$$

Thus, $|L_n(\boldsymbol{\beta}_{S_0}) - L_n(\boldsymbol{\beta})| \leq \|\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta}\|_1[Cr_n + r_n\|F_{n(S_0)}(\boldsymbol{\beta}_1)\| + C_1\|\boldsymbol{\beta}_1 - \widehat{\boldsymbol{\beta}}_{S_0}\|]$ in probability with some constants C and C_1.

On the other hand, $\sum_{j\notin S_0} P_n(|\beta_j|) = \sum_{j\notin S_0}|\beta_j|P_n'(\xi|\beta_j|) \geq P_n'(r_n)\sum_{j\notin S_0}|\beta_j|$ where $\xi \in (0,1)$ by the mean value theorem and the monotonicity of P_n'. Let r_n be small so that $P_n'(r_n) \geq P_n'(0^+)/2$. Hence, $\sum_{j\notin S_0} P_n(|\beta_j|) \geq (P_n'(0^+)/2)\|\boldsymbol{\beta}_{S_0} - \boldsymbol{\beta}\|_1$ in probability. Let r_n be further smaller so that $r_n < P_n'(0^+)/4C$. Then the assertion holds in probability approaching unity for large n.

F Additional simulation results

This subsection provides additional simulation results to compare the performance of the proposed estimator with the post-processed spline estimator of Radchenko (2015) through

Table 1: Average absolute estimation errors (with standard errors in parenthesis), average number of false positives (FP) and average number of false negatives (FN) for DGPs 6 and 7, $k = 3$.

	DGP 6						DGP 7					
Method	σ	n	Estimation error		FP	FN	σ	n	Estimation error		FP	FN
Spline	0.2	70	0.030	(0.010)	0.73	0.02	0.2	70	0.293	(0.044)	1.48	0.68
Hermite			0.031	(0.005)	0.05	0			0.162	(0.062)	0.34	0
Spline		140	0.010	(0.001)	0.24	0		140	0.098	(0.010)	1.05	0.05
Hermite			0.009	(0.004)	0.02	0			0.127	(0.048)	0.13	0
Spline	1	70	0.163	(0.012)	1.55	0.37	0.4	70	0.662	(0.068)	1.84	1.64
Hermite			0.149	(0.044)	0.24	0.24			0.306	(0.121)	0.69	0
Spline		140	0.061	(0.005)	0.49	0.01		140	0.200	(0.012)	0.85	0.57
Hermite			0.041	(0.028)	0.10	0			0.216	(0.088)	0.50	0

the following two designs as used in his study.

$$DGP \quad 6: \qquad y_t = (\mathbf{x}_t' \boldsymbol{\theta}_0)^2 + \sigma e_t,$$

$$DGP \quad 7: \qquad y_t = sin(\mathbf{x}_t' \boldsymbol{\theta}_0 \pi/2) + \sigma e_t,$$

where $\theta_0 = (8, 4, 2, 1, 0, \ldots, 0)'/\sqrt{85}$ and $\mathbf{x}_t \sim N(2\sqrt{17}/5, I_d)$, for $n = 70, 140$ and $d = 100$. The error terms e_t's are independently and identically generated from the standard normal distribution, and two choices of σ are considered for each choice of n. Note that the last nonzero element in the single index coefficient vector is very close to zero, compared to the other nonzero elements. This serves as a scenario in which the signal strength is a bit weak. Following the accuracy measures as in Radchenko (2015), we compute the estimation error (the sum of the absolute differences between the estimated and the true vector) of the index coefficients, the average number of false positives (number of noise predictors identified as signal) and the average number of false negatives (number of signal predictors identified as noise), which are reported in Table 1. The results for the spline estimator are taken directly from Radchenko (2015). The variable selection results for the Hermite polynomial estimator are obtained by applying the screening procedure with the BIC selection, followed by the SCAD penalization. It is observed that our variable selection procedure leads to relatively lower number of false negatives and false positives, while the estimation errors for the two methods are rather comparable for both designs. Overall, the performance of both methods tends to improve as the sample size increases.

References

Belloni, A., Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, 186:345–366.

Dong, C., Gao, J., and Peng, B. (2015). Semiparametric single-index panel data models with cross-sectional dependence. *Journal of Econometrics*, 188:301–312.

Dong, C., Gao, J., and Tjøstheim, D. (2016). Estimation for single-index and partially linear single-index integrated models. *Annals of Statistics*, 44:425–453.

Fan, J. and Liao, Y. (2014). Endogeneity in high dimensions. *Annals of Statistics*, 42:872–917.

Gautschi, W. (2004). *Orthogonal Polynomials: Computation and Approximation*. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford.

Radchenko, P. (2015). High dimensional single index models. *Journal of Multivariate Analysis*, 139:266–282.

Szego, G. (1975). *Orthogonal Polynomials*. Colloquium publications XXIII. American Mathematical Association, Providence, Rhode Island.