Supplemental Document of "*Semiparametric Ultra-High Dimensional Model Averaging of Nonlinear Dynamic Time Series*"

Jia Chen* Degui Li† Oliver Linton‡ Zudi Lu§

Version: February 22, 2017

In this supplemental document, we give some technical lemmas and their proofs in Appendix A, and the proofs of the main asymptotic theorems in Appendix B. An iterative KSIS+PMAMAR procedure, aimed at reducing the false positive rate and potentially increasing the true positive rate, is detailed in Appendix C. The computation times of the methods considered in Examples 5.1 and 5.2 as well as the estimated models for the empirical Example 5.3 are given in Appendix D.

Appendix A: Some technical lemmas

In this appendix, we give some technical lemmas which will be used in the proofs of the main results. The first result is a well-known exponential inequality for the α-mixing sequence which can be found in some existing literature such as Bosq (1998).

Lemma 1. *Let $\{Z_t\}$ be a zero-mean α-mixing process satisfying $\mathsf{P}(|Z_t| \leq B) = 1$ for all $t \geq 1$. Then for each integer $q \in [1, n/2]$ and each $\epsilon > 0$, we have*

$$\mathsf{P}\Big(\Big|\sum_{t=1}^{n} Z_t\Big| > n\epsilon\Big) \leq 4\exp\Big(-\frac{\epsilon^2 q}{8v^2(q)}\Big) + 22\Big(1+\frac{4B}{\epsilon}\Big)^{1/2} q\alpha(\lfloor p \rfloor), \tag{A.1}$$

*Department of Economics and Related Studies, University of York, YO10 5DD, UK. E-mail: jia.chen@york.ac.uk
†Department of Mathematics, University of York, YO10 5DD, UK. E-mail: degui.li@york.ac.uk.
‡Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, CB3 9DD, UK. E-mail: obl20@cam.ac.uk.
§Statistical Sciences Research Institute and School of Mathematical Sciences, University of Southampton, SO17 1BJ, UK. E-mail: Z.Lu@soton.ac.uk.

where $v^2(q) = 2\sigma^2(q)/p^2 + B\epsilon/2$, $p = n/(2q)$,

$$\sigma^2(q) = \max_{1\le j\le 2q-1} \mathsf{E}\ \Big\{(\lfloor jp\rfloor + 1 - jp)Z_{\lfloor jp\rfloor+1} + Z_{\lfloor jp\rfloor+2} + \ldots + Z_{\lfloor (j+1)p\rfloor} + ((j+1)p - \lfloor (j+1)p\rfloor)Z_{\lfloor (j+1)p\rfloor+1}\Big\}^2$$

$\alpha(\cdot)$ *is the* α*-mixing coefficient, and* $\lfloor\cdot\rfloor$ *denotes the integer part.*

Define

$$\mathcal{G}_{ji} = \Big\{\sup_{x_j\in\mathcal{C}_j}\Big|\sum_{t=1}^{n}\Big\{\big(\frac{X_{tj}-x_j}{h_1}\big)^i K\big(\frac{X_{tj}-x_j}{h_1}\big) - \mathsf{E}\big[\big(\frac{X_{tj}-x_j}{h_1}\big)^i K\big(\frac{X_{tj}-x_j}{h_1}\big)\big]\Big\}\Big| \ge (nh_1)^{\frac{1}{2}+\kappa_2}\Big\}, \quad \text{(A.2)}$$

where $0 < \kappa_2 < 1/2$ and $i = 0, 1, \ldots$. The following Lemma gives an upper bound of the probability of the event $\mathcal{G}_{j0}\cup\mathcal{G}_{j1}\cup\mathcal{G}_{j2}$.

Lemma 2. *Suppose that the conditions A1–A3 in Section 3.1 are satisfied. Then we have*

$$\mathsf{P}\big(\mathcal{G}_{j0}\cup\mathcal{G}_{j1}\cup\mathcal{G}_{j2}\big) \le M_1 n^{(5+9\theta_1-2\kappa_2+2\kappa_2\theta_1)/4}\Big[\exp\big(-c_1 n^{2(1-\theta_1)\kappa_2}\big) + \exp\Big(-c_2 n^{(1-\theta_1)(\frac{1}{2}-\kappa_2)}\Big)\Big] \quad \text{(A.3)}$$

for $j = 1, 2, \ldots, p_n + d_n$, *where* c_1, c_2 *and* M_1 *are positive constants which are independent of* j, *and* θ_1 *is defined in the condition A3.*

Proof. We next only prove that

$$\mathsf{P}\big(\mathcal{G}_{j0}\big) \le \frac{M_1}{3}\Big\{n^{(5+9\theta_1-2\kappa_2+2\kappa_2\theta_1)/4}\Big[\exp\big(-c_1 n^{2(1-\theta_1)\kappa_2}\big) + \exp\Big(-c_2 n^{(1-\theta_1)(\frac{1}{2}-\kappa_2)}\Big)\Big]\Big\}, \quad \text{(A.4)}$$

as the same conclusion also holds for $\mathcal{G}_{j1}$ and $\mathcal{G}_{j2}$ with a similar proof. Then the proof of (A.3) can be completed. We cover the uniformly bounded set $\mathcal{C}_j$ by a finite number of intervals $\mathcal{C}_j(k)$ with centre $s_j(k)$ and radius $h_1(nh_1)^{\frac{1}{2}+\kappa_2}/(3c_K n)$, where c_K is a positive constant such that $|K(u) - K(v)| \le c_K|u - v|$. Letting $N_n(j)$ be the total number of $\mathcal{C}_j(k)$, by the condition A2, it is easy to see that

$$\max_{1\le j\le p_n+d_n} N_n(j) = O\big(nh_1^{-1}(nh_1)^{-\frac{1}{2}-\kappa_2}\big). \quad \text{(A.5)}$$

Note that

$$
\begin{aligned}
& \sup_{x_j\in\mathcal{C}_j}\Big|\sum_{t=1}^n\Big\{K\big(\frac{X_{tj}-x_j}{h_1}\big)-\mathsf{E}\big[K\big(\frac{X_{tj}-x_j}{h_1}\big)\big]\Big\}\Big| \\
\leq\ & \max_{1\leq k\leq N_n(j)}\Big|\sum_{t=1}^n\Big\{K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)-\mathsf{E}\big[K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)\big]\Big\}\Big|+ \\
& \max_{1\leq k\leq N_n(j)}\sup_{x_j\in\mathcal{C}_j(k)}\Big|\sum_{t=1}^n\Big[K\big(\frac{X_{tj}-x_j}{h_1}\big)-K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)\Big]\Big|+ \\
& \max_{1\leq k\leq N_n(j)}\sup_{x_j\in\mathcal{C}_j(k)}\Big|\sum_{t=1}^n\mathsf{E}\Big[K\big(\frac{X_{tj}-x_j}{h_1}\big)-K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)\Big]\Big| \\
\leq\ & \max_{1\leq k\leq N_n(j)}\Big|\sum_{t=1}^n\Big\{K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)-\mathsf{E}\big[K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)\big]\Big\}\Big|+\frac{2}{3}(nh_1)^{\frac{1}{2}+\kappa_2},
\end{aligned}
$$

which indicates that

$$
\begin{aligned}
\mathsf{P}\big(\mathcal{G}_{j0}\big) &\leq \mathsf{P}\Big(\max_{1\leq k\leq N_n(j)}\Big|\sum_{t=1}^n\Big\{K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)-\mathsf{E}\big[K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)\big]\Big\}\Big|>\frac{1}{3}(nh_1)^{\frac{1}{2}+\kappa_2}\Big) \\
&\leq \sum_{k=1}^{N_n(j)}\mathsf{P}\Big(\Big|\sum_{t=1}^n\Big\{K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)-\mathsf{E}\big[K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)\big]\Big\}\Big|>\frac{1}{3}(nh_1)^{\frac{1}{2}+\kappa_2}\Big). \qquad \text{(A.6)}
\end{aligned}
$$

Then, by taking $Z_t=K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)-\mathsf{E}\big[K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)\big]$, $B=2\sup_u K(u)$, $p=(nh_1)^{\frac{1}{2}-\kappa_2}$ and $\epsilon=(nh_1)^{\frac{1}{2}+\kappa_2}/(3n)$ in Lemma 1 and noting that $h_1\sim n^{-\theta_1}$, we may show that

$$
\begin{aligned}
& \mathsf{P}\Big(\Big|\sum_{t=1}^n\Big\{K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)-\mathsf{E}\big[K\big(\frac{X_{tj}-s_j(k)}{h_1}\big)\big]\Big\}\Big|>\frac{1}{3}(nh_1)^{\frac{1}{2}+\kappa_2}\Big) \\
\leq\ & 4\exp\big\{-c_1(nh_1)^{2\kappa_2}\big\}+c_3n^{3/2}(nh_1)^{(2\kappa_2-3)/4}\exp\Big\{-c_2(nh_1)^{\frac{1}{2}-\kappa_2}\Big\} \\
=\ & 4\exp\big\{-c_1n^{2(1-\theta_1)\kappa_2}\big\}+c_3n^{[2\kappa_2+3-(2\kappa_2-3)\theta_1]/4}\exp\Big\{-c_2n^{(1-\theta_1)(\frac{1}{2}-\kappa_2)}\Big\}, \qquad \text{(A.7)}
\end{aligned}
$$

where c_1, c_2 and c_3 are positive constants which are independent of j. Combining (A.5)–(A.7), we can prove (A.4), completing the proof of Lemma 2. ■

Lemma 3. *Let $\eta_{tj}=Y_t-m_j(X_{tj})$. Suppose that the conditions A1–A5 are satisfied. Then we have for any $\xi>0$ and $j=1,2,\dots,p_n+d_n$,*

$$
\begin{aligned}
& \mathsf{P}\Big(\sup_{x_j\in\mathcal{C}_j}\Big|\sum_{t=1}^n\eta_{tj}K\big(\frac{X_{tj}-x_j}{h_1}\big)\Big|>\xi(nh_1)n^{-\kappa_1}\Big) \\
\leq\ & M_2n^{1+\frac{7\kappa_1}{4}+\frac{5\theta_1}{2}}\left[\exp\left(-c_4n^{1-2\kappa_1-\theta_1}\right)+\exp\left(-c_5n^{\kappa_1/2}\right)\right], \qquad \text{(A.8)}
\end{aligned}
$$

where $0 < \kappa_1 < (1-\theta_1)/2$, c_4, c_5 and M_2 are positive constants which are independent of j.

Proof. As $\mathsf{E}[\exp\{\varsigma|Y_t|\}] < \infty$ assumed in the condition A5, we may show that

$$\mathsf{E}[\exp\{\varsigma|\eta_{tj}|\}] \le \mathsf{E}[\exp\{\varsigma|Y_t| + \varsigma|m_j(X_{tj})|\}] \le e^{\varsigma c_m}\mathsf{E}[\exp\{\varsigma|Y_t|\}] < \infty. \tag{A.9}$$

Let

$$\nu_n = n^{\kappa_1/2}, \quad \overline{\eta}_{tj} = \eta_{tj} I\big(|\eta_{tj}| \le \nu_n\big), \quad \widetilde{\eta}_{tj} = \eta_{tj} I\big(|\eta_{tj}| > \nu_n\big).$$

As $\mathsf{E}[\eta_{tj}] = 0$, it is easy to show that

$$\eta_{tj} = \eta_{tj} - \mathsf{E}[\eta_{tj}] = \overline{\eta}_{tj} - \mathsf{E}[\overline{\eta}_{tj}] + \widetilde{\eta}_{tj} - \mathsf{E}[\widetilde{\eta}_{tj}].$$

Hence, we have

$$\sum_{t=1}^n \eta_{tj} K\big(\frac{X_{tj}-x_j}{h_1}\big) = \sum_{t=1}^n \big\{\overline{\eta}_{tj} - \mathsf{E}[\overline{\eta}_{tj}]\big\} K\big(\frac{X_{tj}-x_j}{h_1}\big) + \sum_{t=1}^n \big\{\widetilde{\eta}_{tj} - \mathsf{E}[\widetilde{\eta}_{tj}]\big\} K\big(\frac{X_{tj}-x_j}{h_1}\big). \tag{A.10}$$

For sufficiently large k, by (A.9) and the choice of ν_n, we can prove that

$$\mathsf{E}[|\widetilde{\eta}_{tj}|] = \mathsf{E}[|\eta_{tj}|\mathsf{I}(|\eta_{tj}| > \nu_n)] \le \mathsf{E}[|\eta_{tj}|^{k+1}\nu_n^{-k}] = O(\nu_n^{-k}) = o(h_1 n^{-\kappa_1}).$$

Then, we can show that

$$\begin{aligned}
& \mathsf{P}\Big(\sup_{x_j\in\mathcal{C}_j} \Big|\sum_{t=1}^n \big\{\widetilde{\eta}_{tj} - \mathsf{E}[\widetilde{\eta}_{tj}]\big\} K\big(\frac{X_{tj}-x_j}{h_1}\big)\Big| > \frac{1}{2}\xi(nh_1)n^{-\kappa_1}\Big) \\
\le\ & \mathsf{P}\Big(\sup_{x_j\in\mathcal{C}_j} \Big|\sum_{t=1}^n \widetilde{\eta}_{tj} K\big(\frac{X_{tj}-x_j}{h_1}\big)\Big| > \frac{1}{4}\xi(nh_1)n^{-\kappa_1}\Big) \\
\le\ & \mathsf{P}\Big(\max_{1\le t\le n} |\eta_{tj}| > \nu_n\Big) \le \sum_{t=1}^n \mathsf{P}\Big(|\eta_{tj}| > \nu_n\Big) \\
\le\ & n\frac{\mathsf{E}[\exp\{\varsigma|\eta_{tj}|\}]}{\exp\{\varsigma\nu_n\}} = M_2(n\exp\{-\varsigma n^{\kappa_1/2}\})/2,
\end{aligned} \tag{A.11}$$

where M_2 is a sufficiently large positive constant which is independent of j.

We next consider the upper bound for the probability of the event:

$$\Big\{\sup_{x_j\in\mathcal{C}_j} \Big|\sum_{t=1}^n \big\{\overline{\eta}_{tj} - \mathsf{E}[\overline{\eta}_{tj}]\big\} K\big(\frac{X_{tj}-x_j}{h_1}\big)\Big| > \frac{1}{2}\xi(nh_1)n^{-\kappa_1}\Big\}.$$

The argument is similar to the proof of (A.4). We cover $\mathcal{C}_j$ by a finite number of intervals $\mathcal{C}_j^*(k)$ with centre $s_j^*(k)$ and radius $\xi h_1^2 n^{-\kappa_1}/(6c_K\nu_n)$, where c_K is defined as in the proof of Lemma 2. Letting

$N_n^*(j)$ be the total number of $\mathcal{C}_j^*(k)$, the order of $N_n^*(j)$ is $O\big(n^{\kappa_1}h_1^{-2}\nu_n\big)$ uniformly over j. By some standard arguments, we have

$$\begin{aligned}
& \sup_{x_j\in\mathcal{C}_j}\Big|\sum_{t=1}^n\{\overline{\eta}_{tj}-\mathsf{E}[\overline{\eta}_{tj}]\}K\big(\frac{X_{tj}-x_j}{h_1}\big)\Big| \\
\leq\ & \max_{1\leq k\leq N_n^*(j)}\Big|\sum_{t=1}^n\{\overline{\eta}_{tj}-\mathsf{E}[\overline{\eta}_{tj}]\}K\big(\frac{X_{tj}-s_j^*(k)}{h_1}\big)\Big| + \\
& \max_{1\leq k\leq N_n^*(j)}\sup_{x_j\in\mathcal{C}_j^*(k)}\Big|\sum_{t=1}^n\{\overline{\eta}_{tj}-\mathsf{E}[\overline{\eta}_{tj}]\}\Big[K\big(\frac{X_{tj}-x_j}{h_1}\big)-K\big(\frac{X_{tj}-s_j^*(k)}{h_1}\big)\Big]\Big| \\
\leq\ & \max_{1\leq k\leq N_n^*(j)}\Big|\sum_{t=1}^n\{\overline{\eta}_{tj}-\mathsf{E}[\overline{\eta}_{tj}]\}K\big(\frac{X_{tj}-s_j^*(k)}{h_1}\big)\Big| + \frac{1}{3}\xi(nh_1)n^{-\kappa_1}.
\end{aligned}$$

Hence, we have

$$\begin{aligned}
& \mathsf{P}\Big(\sup_{x_j\in\mathcal{C}_j}\Big|\sum_{t=1}^n\{\overline{\eta}_{tj}-\mathsf{E}[\overline{\eta}_{tj}]\}K\big(\frac{X_{tj}-x_j}{h_1}\big)\Big| > \frac{1}{2}\xi(nh_1)n^{-\kappa_1}\Big) \\
\leq\ & \mathsf{P}\Big(\max_{1\leq k\leq N_n^*(j)}\Big|\sum_{t=1}^n\{\overline{\eta}_{tj}-\mathsf{E}[\overline{\eta}_{tj}]\}K\big(\frac{X_{tj}-s_j^*(k)}{h_1}\big)\Big| > \frac{1}{6}\xi(nh_1)n^{-\kappa_1}\Big) \\
\leq\ & \sum_{k=1}^{N_n^*(j)}\mathsf{P}\Big(\Big|\sum_{t=1}^n\{\overline{\eta}_{tj}-\mathsf{E}[\overline{\eta}_{tj}]\}K\big(\frac{X_{tj}-s_j^*(k)}{h_1}\big)\Big| > \frac{1}{6}\xi(nh_1)n^{-\kappa_1}\Big).
\end{aligned} \tag{A.12}$$

Then, by taking $Z_t=\big\{\overline{\eta}_{tj}-\mathsf{E}[\overline{\eta}_{tj}]\big\}K\big(\frac{X_{tj}-s_j^*(k)}{h_1}\big)$, $B=2\nu_n\sup_u K(u)$, $p=n^{\kappa_1}/\nu_n=n^{\kappa_1/2}$ and $\epsilon=\xi h_1 n^{-\kappa_1}/6$ in Lemma 1 and noting that $h_1\sim n^{-\theta_1}$, we can show that

$$\begin{aligned}
& \mathsf{P}\Big(\Big|\sum_{t=1}^n\{\overline{\eta}_{tj}-\mathsf{E}[\overline{\eta}_{tj}]\}K\big(\frac{X_{tj}-s_j^*(k)}{h_1}\big)\Big| > \frac{1}{6}\xi(nh_1)n^{-\kappa_1}\Big) \\
\leq\ & 4\exp\big\{-c_4 n^{1-2\kappa_1}h_1\big\} + c_6 n^{1+\frac{\kappa_1}{4}}h_1^{-1/2}\exp\big\{-c_5 n^{\kappa_1/2}\big\} \\
=\ & 4\exp\big\{-c_4 n^{1-2\kappa_1-\theta_1}\big\} + c_6 n^{1+\frac{\kappa_1}{4}+\frac{\theta_1}{2}}\exp\big\{-c_5 n^{\kappa_1/2}\big\},
\end{aligned} \tag{A.13}$$

where c_4, c_5 and c_6 are positive constants which are independent of j. By (A.12), (A.13) and the definition of $N_n^*(j)$, we can prove that

$$\begin{aligned}
& \mathsf{P}\Big(\sup_{x_j\in\mathcal{C}_j}\Big|\sum_{t=1}^n\{\overline{\eta}_{tj}-\mathsf{E}[\overline{\eta}_{tj}]\}K\big(\frac{X_{tj}-x_j}{h_1}\big)\Big| > \frac{1}{2}\xi(nh_1)n^{-\kappa_1}\Big) \\
\leq\ & \frac{M_2}{2}n^{1+\frac{7\kappa_1}{4}+\frac{5\theta_1}{2}}\big[\exp\big\{-c_4 n^{1-2\kappa_1-\theta_1}\big\}+\exp\big\{-c_5 n^{\kappa_1/2}\big\}\big].
\end{aligned} \tag{A.14}$$

We can complete the proof of (A.8) by using (A.10), (A.11) and (A.14). ■

We next derive an upper bound for the probability of the event

$$\left\{\sup_{x_j\in\mathcal{C}_j}|\hat{m}_j(x_j)-m_j(x_j)|>\xi n^{-2(1-\theta_1)/5}\right\}$$

for any $\xi>0$, where Lemmas 2 and 3 will play a crucial role.

Lemma 4. *Suppose that the conditions A1–A5 in Section 3.1 are satisfied. Then we have for any $\xi>0$ and $j=1,2,\ldots,p_n+d_n$,*

$$\mathsf{P}\Big(\sup_{x_j\in\mathcal{C}_j}|\hat{m}_j(x_j)-m_j(x_j)|>\xi n^{-2(1-\theta_1)/5}\Big)\leq M_1^*(n)+M_2^*(n), \tag{A.15}$$

where

$$M_1^*(n)=2M_1n^{(7+14\theta_1)/6}\exp\left\{-c_7n^{(1-\theta_1)/3}\right\},\quad M_2^*(n)=2M_2n^{(17+18\theta_1)/10}\exp\left\{-c_8n^{(1-\theta_1)/5}\right\},$$

$c_7=\min(c_1,c_2)$, $c_8=\min(c_4,c_5)$, M_1, c_1 and c_2 are defined in Lemma 2, and M_2, c_4 and c_5 are defined in Lemma 3.

Proof. Let $\mathcal{G}_j=\mathcal{G}_{j0}\cup\mathcal{G}_{j1}\cup\mathcal{G}_{j2}$ and the complement $\mathcal{G}_j^c=\mathcal{G}_{j0}^c\cap\mathcal{G}_{j1}^c\cap\mathcal{G}_{j2}^c$. Notice that

$$\begin{aligned}
&\mathsf{P}\Big(\sup_{x_j\in\mathcal{C}_j}|\hat{m}_j(x_j)-m_j(x_j)|>\xi n^{-\kappa_1}\Big)\\
\leq\ &\mathsf{P}\big(\sup_{x_j\in\mathcal{C}_j}|\hat{m}_j(x_j)-m_j(x_j)|>\xi n^{-\kappa_1},\mathcal{G}_j^c\big)+\mathsf{P}\big(\mathcal{G}_j\big).
\end{aligned}\tag{A.16}$$

By Lemma 2 with $\kappa_2=1/6$, we may show that

$$\mathsf{P}\big(\mathcal{G}_j\big)\leq 2M_1n^{(7+14\theta_1)/6}\exp\left\{-c_7n^{(1-\theta_1)/3}\right\}=:M_1^*(n). \tag{A.17}$$

Consider the decomposition:

$$\begin{aligned}
\hat{m}_j(x_j)-m_j(x_j)&=\frac{\sum_{t=1}^n\big[Y_t-m_j(x_j)\big]K\big(\frac{X_{tj}-x_j}{h_1}\big)}{\sum_{t=1}^nK\big(\frac{X_{tj}-x_j}{h_1}\big)}\\
&=\frac{\sum_{t=1}^n\big[Y_t-m_j(X_{tj})\big]K\big(\frac{X_{tj}-x_j}{h_1}\big)}{\sum_{t=1}^nK\big(\frac{X_{tj}-x_j}{h_1}\big)}+\frac{\sum_{t=1}^n\big[m_j(X_{tj})-m_j(x_j)\big]K\big(\frac{X_{tj}-x_j}{h_1}\big)}{\sum_{t=1}^nK\big(\frac{X_{tj}-x_j}{h_1}\big)}\\
&=:I_{n1}(x_j)+I_{n2}(x_j).
\end{aligned}\tag{A.18}$$

By the condition A4 and Taylor's expansion for $m_j(\cdot)$, we have

$$m_j(X_{tj})-m_j(x_j)=m'(x_j)(X_{tj}-x_j)+\frac{1}{2}m_j''(x_{tj})(X_{tj}-x_j)^2,$$

where x_{tj} lies between X_{tj} and x_j. Hence, for $I_{n2}(x_j)$, we have

$$I_{n2}(x_j) \quad = \quad m'(x_j)\frac{\sum_{t=1}^n (X_{tj}-x_j)K\big(\frac{X_{tj}-x_j}{h_1}\big)}{\sum_{t=1}^n K\big(\frac{X_{tj}-x_j}{h_1}\big)} + \frac{1}{2}\cdot\frac{\sum_{t=1}^n m_j''(x_{tj})(X_{tj}-x_j)^2 K\big(\frac{X_{tj}-x_j}{h_1}\big)}{\sum_{t=1}^n K\big(\frac{X_{tj}-x_j}{h_1}\big)}.$$

On the event $\mathcal{G}_j^c$ with $\kappa_2 = 1/6$, as $\theta_1 > 1/6$ and choosing κ_1 as $2(1-\theta_1)/5$,

$$I_{n2}(x_j) = O\big(h_1^2 + (nh_1)^{-1/3}h_1\big) = o(n^{-\kappa_1}). \tag{A.19}$$

Hence, we have

$$\mathsf{P}\Big(\sup_{x_j\in\mathcal{C}_j}|\hat{m}_j(x_j)-m_j(x_j)| > \xi n^{-\kappa_1}, \mathcal{G}_j^c\Big) \le \mathsf{P}\Big(\sup_{x_j\in\mathcal{C}_j}\Big|\sum_{t=1}^n \eta_{tj}K\big(\frac{X_{tj}-x_j}{h_1}\big)\Big| > \xi_1(nh_1)n^{-\kappa_1}\Big), \tag{A.20}$$

where $\xi_1 = \frac{1}{2}\xi\cdot\inf_j\inf_{x_j\in\mathcal{C}_j} f_j(x_j)$. By Lemma 3 with $\kappa_1 = 2(1-\theta_1)/5$, we have

$$\mathsf{P}\Big(\sup_{x_j\in\mathcal{C}_j}\Big|\sum_{t=1}^n \eta_{tj}K\big(\frac{X_{tj}-x_j}{h_1}\big)\Big| > \xi_1(nh_1)n^{-\kappa_1}\Big) \le 2M_2 n^{(17+18\theta_1)/10}\exp\left\{-c_8 n^{(1-\theta_1)/5}\right\},$$

with $c_8 = \min(c_4, c_5)$, which indicates that

$$\mathsf{P}\Big(\sup_{x_j\in\mathcal{C}_j}|\hat{m}_j(x_j)-m_j(x_j)| > \xi n^{-\kappa_1}, \mathcal{G}_j^c\Big) \le 2M_2 n^{(17+18\theta_1)/10}\exp\left\{-c_8 n^{(1-\theta_1)/5}\right\} =: M_2^*(n). \tag{A.21}$$

We can complete the proof of (A.15) by combining (A.16), (A.17) and (A.21). ■

Lemma 5. *Suppose that the conditions B1–B4 and (3.11) in Section 3.2 are satisfied. Then, we have*

$$\left\|\boldsymbol{d}_n(\hat{\mathcal{F}}_n)\right\| := \left\|\mathsf{vec}\left(\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{F}_n^0\right)/\sqrt{n}\right\| = o_P(1), \tag{A.22}$$

where $\mathsf{vec}(\cdot)$ *denotes the vectorization of a matrix, and*

$$\mathcal{M}_{\hat{\mathcal{F}}_n} = I_n - \hat{\mathcal{F}}_n\left(\hat{\mathcal{F}}_n^{\intercal}\hat{\mathcal{F}}_n\right)^{-1}\hat{\mathcal{F}}_n^{\intercal} = I_n - \frac{1}{n}\hat{\mathcal{F}}_n\hat{\mathcal{F}}_n^{\intercal}$$

using the restriction of $\hat{\mathcal{F}}_n^{\intercal}\hat{\mathcal{F}}_n/n = I_r$

Proof. It is easy to see that the PCA method is equivalent to the following constrained least squares method:

$$\left(\hat{\mathcal{F}}_n, \hat{\mathbf{B}}_n\right) = \arg\min_{\mathbf{b}_k,\mathbf{f}_t}\sum_{k=1}^{p_n}\sum_{t=1}^n\left(Z_{tk}-\mathbf{b}_k^{\intercal}\mathbf{f}_t\right)^2 = \arg\min_{\mathcal{F}_n,\mathbf{B}_n}\left\|\mathcal{Z}_n - \mathcal{F}_n\mathbf{B}_n^{\intercal}\right\|_F^2, \tag{A.23}$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix, and the $n\times r$ matrix $\mathcal{F}_n$ and the $p_n\times r$ matrix $\mathbf{B}_n$

need to satisfy

$$\frac{1}{n}\mathcal{F}_n^\intercal\mathcal{F}_n = I_r, \quad \frac{1}{p_n}\mathbf{B}_n^\intercal\mathbf{B}_n \text{ is diagonal.} \tag{A.24}$$

Denote $\mathcal{M}_{\mathcal{F}_n} = I_n - \mathcal{F}_n\left(\mathcal{F}_n^\intercal\mathcal{F}_n\right)^{-1}\mathcal{F}_n^\intercal =: I_n - \mathcal{P}_{\mathcal{F}_n}$ and $\mathcal{L}_n(\mathcal{F}_n) = \mathsf{Tr}\left(\mathcal{Z}_n^\intercal\mathcal{M}_{\mathcal{F}_n}\mathcal{Z}_n\right)$, where $\mathsf{Tr}(\cdot)$ is the trace of a square matrix.

From (A.23), we may show that

$$\mathcal{L}_n(\hat{\mathcal{F}}_n) - \mathcal{L}_n(\mathcal{F}_n^0) = \mathsf{Tr}\left(\mathcal{Z}_n^\intercal\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{Z}_n\right) - \mathsf{Tr}\left(\mathcal{Z}_n^\intercal\mathcal{M}_{\mathcal{F}_n^0}\mathcal{Z}_n\right) \le 0. \tag{A.25}$$

Using the fact that $\mathcal{M}_{\mathcal{F}_n^0}\mathcal{F}_n^0 = \mathbf{0}$ and $(\mathcal{F}_n^0)^\intercal\mathcal{M}_{\mathcal{F}_n^0} = \mathbf{0}$, and then by (2.13), we have

$$\begin{aligned}
\mathcal{L}_n(\hat{\mathcal{F}}_n) - \mathcal{L}_n(\mathcal{F}_n^0) &= \mathsf{Tr}\left(\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{F}_n^0(\mathbf{B}_n^0)^\intercal\mathbf{B}_n^0(\mathcal{F}_n^0)^\intercal\mathcal{M}_{\hat{\mathcal{F}}_n}\right) + \mathsf{Tr}\left(\mathcal{U}_n^\intercal\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{U}_n\right) - \\
&\quad \mathsf{Tr}\left(\mathcal{U}_n^\intercal\mathcal{M}_{\mathcal{F}_n^0}\mathcal{U}_n\right) + \mathsf{Tr}\left(\mathcal{U}_n^\intercal\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{F}_n^0(\mathbf{B}_n^0)^\intercal\right) + \\
&\quad \mathsf{Tr}\left(\mathbf{B}_n^0(\mathcal{F}_n^0)^\intercal\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{U}_n\right) \\
&= \mathsf{Tr}\left(\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{F}_n^0(\mathbf{B}_n^0)^\intercal\mathbf{B}_n^0(\mathcal{F}_n^0)^\intercal\mathcal{M}_{\hat{\mathcal{F}}_n}\right) + \mathsf{Tr}\left(\mathcal{U}_n^\intercal\mathcal{P}_{\hat{\mathcal{F}}_n}\mathcal{U}_n\right) - \\
&\quad \mathsf{Tr}\left(\mathcal{U}_n^\intercal\mathcal{P}_{\mathcal{F}_n^0}\mathcal{U}_n\right) + \mathsf{Tr}\left(\mathcal{U}_n^\intercal\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{F}_n^0(\mathbf{B}_n^0)^\intercal\right) + \\
&\quad \mathsf{Tr}\left(\mathbf{B}_n^0(\mathcal{F}_n^0)^\intercal\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{U}_n\right),
\end{aligned} \tag{A.26}$$

where $\mathcal{U}_n = (\mathbf{U}_1, \dots, \mathbf{U}_n)^\intercal$.

We next prove that the last four terms on the right hand side of (A.26) are $o_P(np_n)$. Start with $\mathsf{Tr}\left(\mathcal{U}_n^\intercal\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{F}_n^0(\mathbf{B}_n^0)^\intercal\right)$. Note that

$$\mathsf{Tr}\left(\mathcal{U}_n^\intercal\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{F}_n^0(\mathbf{B}_n^0)^\intercal\right) = \mathsf{Tr}\left(\mathcal{U}_n^\intercal\mathcal{F}_n^0(\mathbf{B}_n^0)^\intercal\right) - \frac{1}{n}\mathsf{Tr}\left(\mathcal{U}_n^\intercal\hat{\mathcal{F}}_n\hat{\mathcal{F}}_n^\intercal\mathcal{F}_n^0(\mathbf{B}_n^0)^\intercal\right) \tag{A.27}$$

using the restriction of $\hat{\mathcal{F}}_n^\intercal\hat{\mathcal{F}}_n/n = I_r$. By the conditions B2 and B4, the Cauchy-Schwarz inequality and some standard arguments, we may show that

$$\begin{aligned}
\mathsf{Tr}\left(\mathcal{U}_n^\intercal\mathcal{F}_n^0(\mathbf{B}_n^0)^\intercal\right) &= \sum_{t=1}^n\sum_{k=1}^{p_n} u_{tk}(\mathbf{f}_t^0)^\intercal\mathbf{b}_k^0 = \left(\sum_{t=1}^n\left\|\mathbf{f}_t^0\right\|^2\right)^{1/2}\left(\sum_{t=1}^n\left\|\sum_{k=1}^{p_n}u_{tk}\mathbf{b}_k^0\right\|^2\right)^{1/2} \\
&= O_P(n^{1/2})\cdot O_P(n^{3/4}p_n^{1/2}) = O_P(n^{5/4}p_n^{1/2}).
\end{aligned} \tag{A.28}$$

On the other hand, by some similar calculations and using the fact that $\|\hat{\mathcal{F}}_n^\intercal\mathcal{F}_n^0\|_F = O_P(n)$, we can also prove that

$$\frac{1}{n}\mathsf{Tr}\left(\mathcal{U}_n^\intercal\hat{\mathcal{F}}_n\hat{\mathcal{F}}_n^\intercal\mathcal{F}_n^0(\mathbf{B}_n^0)^\intercal\right) = O_P(n^{5/4}p_n^{1/2}). \tag{A.29}$$

By (A.27)–(A.29) and using the condition of $n = o(p_n^2)$, we have

$$\mathsf{Tr}\left(\mathcal{U}_n^\intercal\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{F}_n^0(\mathbf{B}_n^0)^\intercal\right) = O_P(n^{5/4}p_n^{1/2}) = o_P(np_n). \tag{A.30}$$

Analogously, we may also show that

$$\mathsf{Tr}\left(\mathbf{B}_n^0(\mathcal{F}_n^0)^\intercal \mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{U}_n\right) = o_P(np_n). \tag{A.31}$$

We next consider $\mathsf{Tr}\left(\mathcal{U}_n^\intercal \mathcal{P}_{\hat{\mathcal{F}}_n}\mathcal{U}_n\right)$. Note that

$$\begin{aligned}
\mathsf{Tr}\left(\mathcal{U}_n^\intercal \mathcal{P}_{\hat{\mathcal{F}}_n}\mathcal{U}_n\right) &= \frac{1}{n}\sum_{k=1}^{p_n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\mathsf{Tr}(\hat{\mathbf{f}}_{t_1}\hat{\mathbf{f}}_{t_2}^\intercal)u_{t_1k}u_{t_2k} \\
&= \frac{1}{n}\sum_{k=1}^{p_n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\mathsf{Tr}(\hat{\mathbf{f}}_{t_1}\hat{\mathbf{f}}_{t_2}^\intercal)\mathsf{E}\left[u_{t_1k}u_{t_2k}\right] + \\
&\quad \frac{1}{n}\sum_{k=1}^{p_n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\mathsf{Tr}(\hat{\mathbf{f}}_{t_1}\hat{\mathbf{f}}_{t_2}^\intercal)\left(u_{t_1k}u_{t_2k} - \mathsf{E}\left[u_{t_1k}u_{t_2k}\right]\right),
\end{aligned} \tag{A.32}$$

where we again have used the fact of $\hat{\mathcal{F}}_n^\intercal\hat{\mathcal{F}}_n/n = I_r$. For the first term on the right hand side of (A.32), by the conditions B1 and B4, and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\left|\sum_{k=1}^{p_n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\mathsf{Tr}(\hat{\mathbf{f}}_{t_1}\hat{\mathbf{f}}_{t_2}^\intercal)\mathsf{E}\left[u_{t_1k}u_{t_2k}\right]\right| \\
\le\ & Cp_n\left(\sum_{t_1=1}^{n}\left\|\hat{\mathbf{f}}_{t_1}\right\|^2\sum_{t_2=1}^{n}\left\|\hat{\mathbf{f}}_{t_2}\right\|^2\right)^{1/2}\left(\sum_{t_1=1}^{n}\sum_{t_1=1}^{n}\mathsf{E}^2\left[u_{t_1k}u_{t_2k}\right]\right)^{1/2} \\
=\ & p_n\cdot O_P(n)\cdot O_P(n^{1/2}).
\end{aligned} \tag{A.33}$$

For the second term on the right hand side of (A.32), letting $u(t_1,t_2) = \sum_{k=1}^{p_n}\left(u_{t_1k}u_{t_2k} - \mathsf{E}\left[u_{t_1k}u_{t_2k}\right]\right)$, by (3.10) in the condition B4 and the Cauchy-Schwarz inequality again, we have

$$\begin{aligned}
&\left|\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\mathsf{Tr}(\hat{\mathbf{f}}_{t_1}\hat{\mathbf{f}}_{t_2}^\intercal)\sum_{k=1}^{p_n}\left(u_{t_1k}u_{t_2k} - \mathsf{E}\left[u_{t_1k}u_{t_2k}\right]\right)\right| \\
\le\ & \left(\sum_{t_1=1}^{n}\left\|\hat{\mathbf{f}}_{t_1}\right\|^2\sum_{t_2=1}^{n}\left\|\hat{\mathbf{f}}_{t_2}\right\|^2\right)^{1/2}\left(\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}u^2(t_1,t_2)\right)^{1/2} \\
=\ & O_P(n)\cdot O_P(np_n^{1/2}n^{1/4}) = O_P(n^{9/4}p_n^{1/2}).
\end{aligned} \tag{A.34}$$

In view of (A.32)–(A.34), we have

$$\mathsf{Tr}\left(\mathcal{U}_n^\intercal \mathcal{P}_{\hat{\mathcal{F}}_n}\mathcal{U}_n\right) = O_P(n^{5/4}p_n^{1/2} + n^{1/2}p_n) = o_P(np_n). \tag{A.35}$$

We finally consider $\mathsf{Tr}\left(\mathcal{U}_n^\intercal \mathcal{P}_{\mathcal{F}_n^0}\mathcal{U}_n\right)$. By the conditions B1 and B2 as well as the central limit theorem, we have

$$\frac{1}{n}\sum_{t=1}^{n}\mathbf{f}_t^0(\mathbf{f}_t^0)^\intercal - \mathbf{\Lambda}_F = O_P(n^{-1/2}),$$

which indicates that

$$\begin{aligned}
\mathsf{Tr}\left(\mathcal{U}_n^\intercal \mathcal{P}_{\mathcal{F}_n^0}\mathcal{U}_n\right) &= \frac{1}{n}\sum_{k=1}^{p_n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\mathsf{Tr}(\mathbf{f}_{t_1}^0\mathbf{\Lambda}_F^{-1}(\mathbf{f}_{t_2}^0)^\intercal)u_{t_1k}u_{t_2k} + O_P(n^{-3/2})\sum_{k=1}^{p_n}\left\|\sum_{t_1=1}^{n}\mathbf{f}_{t_1}^0 u_{t_1k}\right\|\left\|\sum_{t_2=1}^{n}\mathbf{f}_{t_2}^0 u_{t_2k}\right\| \\
&= \frac{1}{n}\sum_{k=1}^{p_n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\mathsf{Tr}(\mathbf{f}_{t_1}^0\mathbf{\Lambda}_F^{-1}(\mathbf{f}_{t_2}^0)^\intercal)u_{t_1k}u_{t_2k} + O_P(p_n n^{1/2}) \\
&= \frac{1}{n}\sum_{k=1}^{p_n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\mathsf{Tr}(\mathbf{f}_{t_1}^0\mathbf{\Lambda}_F^{-1}(\mathbf{f}_{t_2}^0)^\intercal)u_{t_1k}u_{t_2k} + o_P(np_n), \qquad \text{(A.36)}
\end{aligned}$$

where we have used the result that

$$\max_{1\le k\le p_n}\left\|\sum_{t=1}^{n}\mathbf{f}_t^0 u_{tk}\right\| = O_P(n), \qquad \text{(A.37)}$$

which can be proved by using the exponential inequality in Lemma 1 and the arguments in Lemmas 2 and 3. Following the arguments in the proof of (A.35), we can similarly show that

$$\frac{1}{n}\sum_{k=1}^{p_n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\mathsf{Tr}(\mathbf{f}_{t_1}^0\mathbf{\Lambda}_F^{-1}(\mathbf{f}_{t_2}^0)^\intercal)u_{t_1k}u_{t_2k} = O_P(n^{5/4}p_n^{1/2} + n^{1/2}p_n), \qquad \text{(A.38)}$$

which together with (A.36), implies that

$$\mathsf{Tr}\left(\mathcal{U}_n^\intercal \mathcal{P}_{\mathcal{F}_n^0}\mathcal{U}_n\right) = o_P(np_n). \qquad \text{(A.39)}$$

Hence, by (A.30), (A.31), (A.35) and (A.39), we have

$$\frac{1}{np_n}\left[\mathcal{L}_n(\hat{\mathcal{F}}_n) - \mathcal{L}_n(\mathcal{F}_n^0)\right] = \frac{1}{np_n}\mathsf{Tr}\left(\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{F}_n^0(\mathbf{B}_n^0)^\intercal\mathbf{B}_n^0(\mathcal{F}_n^0)^\intercal\mathcal{M}_{\hat{\mathcal{F}}_n}\right) + o_P(1). \qquad \text{(A.40)}$$

Define

$$\mathbf{\Sigma}_n(\mathbf{B}_n^0) = \frac{1}{p_n}(\mathbf{B}_n^0)^\intercal\mathbf{B}_n^0 \otimes I_n, \quad \boldsymbol{d}_n(\hat{\mathcal{F}}_n) = \mathsf{vec}\left(\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{F}_n^0\right)/\sqrt{n},$$

where $\otimes$ denotes the Kronecker product. It is easy to verify that

$$\frac{1}{np_n}\mathsf{Tr}\left(\mathcal{M}_{\hat{\mathcal{F}}_n}\mathcal{F}_n^0(\mathbf{B}_n^0)^\intercal\mathbf{B}_n^0(\mathcal{F}_n^0)^\intercal\mathcal{M}_{\hat{\mathcal{F}}_n}\right) = \mathbf{d}_n^\intercal(\hat{\mathcal{F}}_n)\mathbf{\Sigma}_n(\mathbf{B}_n^0)\mathbf{d}_n(\hat{\mathcal{F}}_n) + o_P(1). \qquad \text{(A.41)}$$

By the condition B3, the smallest eigenvalue of $\mathbf{\Sigma}_n(\mathbf{B}_n^0)$ is positive and bounded away from zero.

Therefore we can prove that

$$0 \leq \mathbf{d}_n^{\intercal}(\hat{\mathcal{F}}_n)\mathbf{\Sigma}_n(\mathbf{B}_n^0)\mathbf{d}_n(\hat{\mathcal{F}}_n) = o_P(1), \tag{A.42}$$

which leads to (A.22), completing the proof of Lemma 5. ■

Appendix B: Proofs of the main results

In this appendix, we provide the detailed proofs of the asymptotic results given in Section 3.

Proof of Theorem 1(i) By the definition of $\hat{\mathsf{v}}(j)$, we have for $j = 1, 2, \ldots, p_n + d_n$,

$$\begin{aligned}
\hat{\mathsf{v}}(j) - \mathsf{v}(j) &= \frac{1}{n}\sum_{t=1}^{n} \hat{m}_j^2(X_{tj}) - \Big[\frac{1}{n}\sum_{t=1}^{n} \hat{m}_j(X_{tj})\Big]^2 - \mathsf{var}\big(m_j(X_{tj})\big) \\
&= \Big\{\frac{1}{n}\sum_{t=1}^{n} \big[\hat{m}_j^2(X_{tj}) - m_j^2(X_{tj})\big]\Big\} - \Big\{\Big[\frac{1}{n}\sum_{t=1}^{n} \hat{m}_j(X_{tj})\Big]^2 - \Big[\frac{1}{n}\sum_{t=1}^{n} m_j(X_{tj})\Big]^2\Big\} + \\
&\quad \Big\{\frac{1}{n}\sum_{t=1}^{n} \big(m_j^2(X_{tj}) - \mathsf{E}[m_j^2(X_{tj})]\big)\Big\} - \Big\{\Big[\frac{1}{n}\sum_{t=1}^{n} m_j(X_{tj})\Big]^2 - \mathsf{E}^2\big[m_j(X_{tj})\big]\Big\} \\
&=: \Pi_{nj}(1) + \Pi_{nj}(2) + \Pi_{nj}(3) + \Pi_{nj}(4).
\end{aligned} \tag{B.1}$$

For $\Pi_{nj}(1)$, note that

$$\begin{aligned}
\big|\Pi_{nj}(1)\big| &= \frac{1}{n}\sum_{t=1}^{n} \big|\hat{m}_j^2(X_{tj}) - m_j^2(X_{tj})\big| \\
&= \frac{1}{n}\sum_{t=1}^{n} \big[\hat{m}_j(X_{tj}) - m_j(X_{tj})\big]^2 + \frac{2}{n}\sum_{t=1}^{n} \big|m_j(X_{tj})\big|\big|\hat{m}_j(X_{tj}) - m_j(X_{tj})\big| \\
&\leq \sup_{x_j \in \mathcal{C}_j} \big|\hat{m}_j(x_j) - m_j(x_j)\big|^2 + 2c_m \cdot \sup_{x_j \in \mathcal{C}_j} \big|\hat{m}_j(x_j) - m_j(x_j)\big|.
\end{aligned} \tag{B.2}$$

By (B.2) and Lemma 4, we readily obtain

$$\mathsf{P}\Big(\big|\Pi_{nj}(1)\big| > \frac{\delta_1}{4} n^{-2(1-\theta_1)/5}\Big) \leq M_2^{\diamond} n^{(17+18\theta_1)/10} \exp\big\{-c_8 n^{(1-\theta_1)/5}\big\}, \tag{B.3}$$

where $M_2^{\diamond} > 0$ is a sufficiently large constant independent of j and c_8 is defined in Lemma 4. Analogously, we can also show that

$$\mathsf{P}\Big(\big|\Pi_{nj}(2)\big| > \frac{\delta_1}{4} n^{-2(1-\theta_1)/5}\Big) \leq M_2^{\diamond} n^{(17+18\theta_1)/10} \exp\big\{-c_8 n^{(1-\theta_1)/5}\big\}. \tag{B.4}$$

Using Lemma 1 with $Z_t = m_j(X_{tj})$ or $m_j^2(X_{tj})$, $p = \lfloor n^{2(1-\theta_1)/5} \rfloor$ and $\epsilon = n^{-2(1-\theta_1)/5}$, we may show that

$$\mathsf{P}\Big(\big|\Pi_{nj}(3)\big| > \frac{\delta_1}{4} n^{-2(1-\theta_1)/5}\Big) + \mathsf{P}\Big(\big|\Pi_{nj}(4)\big| > \frac{\delta_1}{4} n^{-2(1-\theta_1)/5}\Big) = o\left(n^{(17+18\theta_1)/10} \exp\left\{-c_8 n^{(1-\theta_1)/5}\right\}\right). \tag{B.5}$$

Then, by (B.1) and (B.3)–(B.5), we can prove that

$$\mathsf{P}\Big(\big|\hat{\mathsf{v}}(j) - \mathsf{v}(j)\big| > \delta_1 n^{-2(1-\theta_1)/5}\Big) \leq 3M_2^{\diamond} n^{(17+18\theta_1)/10} \exp\left\{-c_8 n^{(1-\theta_1)/5}\right\}, \tag{B.6}$$

which indicates that

$$\begin{aligned} & \mathsf{P}\Big(\max_{1\leq j\leq p_n+d_n} \big|\hat{\mathsf{v}}(j) - \mathsf{v}(j)\big| > \delta_1 n^{-2(1-\theta_1)/5}\Big) \\ \leq & \sum_{j=1}^{p_n+d_n} \mathsf{P}\Big(\big|\hat{\mathsf{v}}(j) - \mathsf{v}(j)\big| > \delta_1 n^{-2(1-\theta_1)/5}\Big) \\ \leq & \; O\left((p_n+d_n) n^{(17+18\theta_1)/10} \exp\left\{-c_8 n^{(1-\theta_1)/5}\right\}\right). \end{aligned} \tag{B.7}$$

Choosing $M(n) = (p_n + d_n) n^{(17+18\theta_1)/10}$ and $\delta_2 = c_8$, we can complete the proof of Theorem 1(i).

(ii) By the definition of $\hat{\mathcal{S}}$, using the condition that $\min_{j\in\mathcal{S}} \mathsf{v}(j) \geq 2\delta_1 n^{-2(1-\delta_1)/5}$ and following the proof of Theorem 1(i), we have

$$\begin{aligned} \mathsf{P}\big(\mathcal{S} \subset \hat{\mathcal{S}}\big) &= \mathsf{P}\big(\min_{j\in\mathcal{S}} \hat{\mathsf{v}}(j) \geq \rho_n\big) = \mathsf{P}\big(\min_{j\in\mathcal{S}} \hat{\mathsf{v}}(j) \geq \delta_1 n^{-2(1-\theta_1)/5}\big) \\ &= \mathsf{P}\big(\min_{j\in\mathcal{S}} \mathsf{v}(j) - \min_{j\in\mathcal{S}} \hat{\mathsf{v}}(j) \leq \min_{j\in\mathcal{S}} \mathsf{v}(j) - \delta_1 n^{-2(1-\theta_1)/5}\big) \\ &\geq \mathsf{P}\big(\min_{j\in\mathcal{S}} \mathsf{v}(j) - \min_{j\in\mathcal{S}} \hat{\mathsf{v}}(j) \leq 2\delta_1 n^{-2(1-\theta_1)/5} - \delta_1 n^{-2(1-\theta_1)/5}\big) \\ &\geq \mathsf{P}\big(\max_{j\in\mathcal{S}} \big|\hat{\mathsf{v}}(j) - \mathsf{v}(j)\big| \leq \delta_1 n^{-2(1-\theta_1)/5}\big) \\ &= 1 - \mathsf{P}\big(\max_{j\in\mathcal{S}} \big|\hat{\mathsf{v}}(j) - \mathsf{v}(j)\big| > \delta_1 n^{-2(1-\theta_1)/5}\big) \\ &\geq 1 - O\left(M_{\mathcal{S}}(n) \exp\big\{-\delta_2 n^{(1-\theta_1)/5}\big\}\right). \end{aligned} \tag{B.8}$$

Then, we complete the proof of (3.3).

The proof of Theorem 1 has been completed. ■

Proof of Theorem 2(i) Recall that $\mathbf{w}_n = (w_1, \ldots, w_{q_n})^\intercal$ and $\mathbf{w}_o = (w_{o1}, \ldots, w_{oq_n})^\intercal = \left[\mathbf{w}_o^\intercal(1), \mathbf{w}_o^\intercal(2)\right]^\intercal$, where $\mathbf{w}_o(1)$ is composed of non-zero weights with dimension s_n, and $\mathbf{w}_o(2)$ is composed of zero weights with dimension $(q_n - s_n)$. Let $\mathcal{Q}_n(\cdot)$ be defined as in (2.9) and $\epsilon_n = \sqrt{q_n}\left(n^{-1/2} + a_n\right)$. In order to prove the convergence rate in Theorem 2(i), as in Fan and Peng (2004), it suffices to show that there exists a sufficiently large constant $C_\diamond > 0$ such that

$$\lim_{n\to\infty} \mathsf{P}\left(\inf_{\|\mathbf{u}\|=C_\diamond} \mathcal{Q}_n(\mathbf{w}_o+\epsilon_n\mathbf{u}) > \mathcal{Q}_n(\mathbf{w}_o)\right) = 1, \tag{B.9}$$

where $\mathbf{u} = (u_1,\ldots,u_{q_n})^\intercal$. In fact, (B.9) implies that there exists a minimum $\hat{\mathbf{w}}_n$ in the ball $\{\mathbf{w}_o+\epsilon_n\mathbf{u}:\ \|\mathbf{u}\|\le C_\diamond\}$, such that $\|\hat{\mathbf{w}}_n-\mathbf{w}_o\| = O_P(\epsilon_n)$.

Observe that

$$\begin{aligned}
& \mathcal{Q}_n(\mathbf{w}_o+\epsilon_n\mathbf{u}) - \mathcal{Q}_n(\mathbf{w}_o) \\
= & \left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o+\epsilon_n\mathbf{u})\right]^\intercal\left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o+\epsilon_n\mathbf{u})\right] + n\sum_{j=1}^{q_n} p_\lambda(|w_{oj}+\epsilon_n u_j|) \\
& -\left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o)\right]^\intercal\left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o)\right] - n\sum_{j=1}^{q_n} p_\lambda(|w_{oj}|) \\
\ge & \left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o+\epsilon_n\mathbf{u})\right]^\intercal\left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o+\epsilon_n\mathbf{u})\right] - \left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o)\right]^\intercal\left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o)\right] \\
& +n\sum_{j=1}^{s_n} p_\lambda(|w_{oj}+\epsilon_n u_j|) - n\sum_{j=1}^{s_n} p_\lambda(|w_{oj}|) \\
= & \ \Xi_{n1} + \Xi_{n2},
\end{aligned} \tag{B.10}$$

where

$$\begin{aligned}
\Xi_{n1} &= \left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o+\epsilon_n\mathbf{u})\right]^\intercal\left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o+\epsilon_n\mathbf{u})\right] - \left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o)\right]^\intercal\left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o)\right], \\
\Xi_{n2} &= n\sum_{j=1}^{s_n}\left[p_\lambda(|w_{oj}+\epsilon_n u_j|) - p_\lambda(|w_{oj}|)\right].
\end{aligned}$$

By the definition of $\hat{\mathcal{M}}(\cdot)$ in Section 2.1 and some elementary calculations, we have

$$\begin{aligned}
\Xi_{n1} &= -2\epsilon_n\mathbf{u}^\intercal\mathcal{S}_n^\intercal(\mathcal{Y})\left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o)\right] + \epsilon_n^2\mathbf{u}^\intercal\mathcal{S}_n^\intercal(\mathcal{Y})\mathcal{S}_n(\mathcal{Y})\mathbf{u} \\
&=: \Xi_{n1}(1) + \Xi_{n1}(2).
\end{aligned}$$

Following the proof of Theorem 3.3 in Li *et al* (2015), we can show that

$$\left\|\mathcal{S}_n^\intercal(\mathcal{Y})\left[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_o)\right]\right\| = O_P(\sqrt{nq_n}),$$

which indicates that

$$|\Xi_{n1}(1)| = O_P(\epsilon_n\sqrt{nq_n})\cdot\|\mathbf{u}\|. \tag{B.11}$$

We next consider $\Xi_{n1}(2)$. By the definition of $m_j^*(\cdot)$ in Section 2.1 and the uniform consistency result in Theorem 3.1 of Li *et al* (2012), we have, uniformly for x_j and $j = 1, 2, \ldots, q_n$,

$$\hat{m}_j^*(x_j) - m_j^*(x_j) = O_P(\tau_n + h_2^2), \tag{B.12}$$

where τ_n is defined in the condition A7. Observe that

$$\mathcal{S}_n^\intercal(\mathcal{Y})\mathcal{S}_n(\mathcal{Y}) = \mathcal{M}_n^\intercal\mathcal{M}_n + \big(\mathcal{S}_n(\mathcal{Y}) - \mathcal{M}_n\big)^\intercal \mathcal{M}_n + \mathcal{M}_n^\intercal\big(\mathcal{S}_n(\mathcal{Y}) - \mathcal{M}_n\big) + \big(\mathcal{S}_n(\mathcal{Y}) - \mathcal{M}_n\big)^\intercal\big(\mathcal{S}_n(\mathcal{Y}) - \mathcal{M}_n\big), \tag{B.13}$$

where $\mathcal{M}_n = \big[\mathcal{M}(1), \ldots, \mathcal{M}(q_n)\big]$ and $\mathcal{M}(j)$, $j = 1, 2, \ldots, q_n$, are defined in Section 2.1. Note that $q_n = o(n^{1/2})$ by using $n^{\frac{1}{2}-\xi}h_2 \to \infty$ and $q_n^2 h_2^2 = o(1)$ in the condition A7. Then, for any $\xi_* > 0$, by Chebyshev's inequality and following the proof of Lemma 8 in Fan and Peng (2004), we have

$$\mathsf{P}\left(\left\|\frac{1}{n}\mathcal{M}_n^\intercal\mathcal{M}_n - \mathbf{\Lambda}_n\right\|_F > \xi_*\right) \leq \frac{1}{\xi_*^2 n^2}\cdot \mathsf{E}\left[\left\|\mathcal{M}_n^\intercal\mathcal{M}_n - \mathbf{\Lambda}_n\right\|_F^2\right] = O\left(q_n^2/n\right) = o(1).$$

Hence, we have

$$\left\|\frac{1}{n}\mathcal{M}_n^\intercal\mathcal{M}_n - \mathbf{\Lambda}_n\right\|_F = o_P(1). \tag{B.14}$$

Equation (B.14) and the condition A6 imply that $\mathbf{u}^\intercal\left(\mathcal{M}_n^\intercal\mathcal{M}_n/n\right)\mathbf{u}$ is asymptotically dominated by $\mathbf{u}^\intercal\mathbf{\Lambda}_n\mathbf{u}$. As $q_n^2(\tau_n + h_2^2) = o(1)$ in the condition A7, we can easily prove that the Frobenius norm for the last three matrices on the right hand side of (B.13) tends to zero. Hence, we have

$$\Xi_{n1}(2) \geq \frac{\chi n\epsilon_n^2}{2}\cdot\|\mathbf{u}\|^2 \tag{B.15}$$

in probability. By (B.11), (B.15) and taking the constant $C_\diamond$ sufficiently large, $\Xi_{n1}(1)$ would be dominated by $\Xi_{n1}(2)$ asymptotically.

For Ξ_{n2}, by the condition A8 and Taylor's expansion for the penalty function, we have

$$\begin{aligned}\Xi_{n2} &= n\sum_{j=1}^{s_n}\big[p_\lambda(|w_{oj} + \epsilon_n u_j|) - p_\lambda(|w_{oj}|)\big]\\ &= O_P(n\epsilon_n a_n\sqrt{q_n})\cdot\|\mathbf{u}\| + O_P(n\epsilon_n^2 b_n)\cdot\|\mathbf{u}\|^2,\end{aligned} \tag{B.16}$$

where w_{oj}^* lies between w_{oj} and $w_{oj} + \epsilon_n u_j$. By the condition A8 , Ξ_{n2} would be also dominated by $\Xi_{n1}(2)$ by taking the constant $C_\diamond$ sufficiently large. We thus complete the proof of (B.9) in view of (B.10), (B.11), (B.15) and (B.16).

(ii) Let $\hat{\mathbf{w}}_n(1)$ and $\hat{\mathbf{w}}_n(2)$ be the estimators of $\mathbf{w}_o(1)$ and $\mathbf{w}_o(2)$, respectively. To prove Theorem 2(ii), it suffices to show that for any constant c_* and any given $\mathbf{w}_n(1)$ satisfying $\|\mathbf{w}_n(1) - \mathbf{w}_o(1)\| = O_P(\epsilon_n^*)$, we have

$$\mathcal{Q}_n\big(\big[\mathbf{w}_n^\intercal(1), \mathbf{0}^\intercal\big]^\intercal\big) = \min_{\|\mathbf{w}_n(2)\|\leq c_*\epsilon_n^*}\mathcal{Q}_n\big(\big[\mathbf{w}_n^\intercal(1), \mathbf{w}_n^\intercal(2)\big]^\intercal\big), \tag{B.17}$$

where $\epsilon_n^* = \sqrt{q_n/n}$, $\mathbf{w}_n(2)$ is a $(q_n - s_n)$-dimensional vector. By (B.17), Theorem 2(i) and noting that $a_n = O\big(n^{-1/2}\big)$ in the condition A8, it is easy to prove that $\hat{\mathbf{w}}_n(2) = \mathbf{0}$.

As in Fan and Li (2001), to prove (B.17), it is sufficient to show that, with probability approaching one, for any q_n-dimensional vector $\mathbf{w}_n^\intercal = \big[\mathbf{w}_n^\intercal(1), \mathbf{w}_n^\intercal(2)\big]$ with $\mathbf{w}_n(1)$ satisfying $\|\mathbf{w}_n(1) - \mathbf{w}_o(1)\| = O_P(\epsilon_n^*)$ and for $j = s_n + 1, \ldots, q_n$,

$$\frac{\partial \mathcal{Q}_n(\mathbf{w}_n)}{\partial w_j} > 0, \quad 0 < w_j < \epsilon_n^*, \tag{B.18}$$

and

$$\frac{\partial \mathcal{Q}_n(\mathbf{w}_n)}{\partial w_j} < 0, \quad -\epsilon_n^* < w_j < 0, \tag{B.19}$$

where $\mathbf{w}_n^\intercal(2) = (w_{s_n+1}, \ldots, w_{d_n})$.

Note that

$$\frac{\partial \mathcal{Q}_n(\mathbf{w}_n)}{\partial w_j} = \frac{\partial \mathcal{L}_n(\mathbf{w}_n)}{\partial w_j} + np_\lambda'(|w_j|)\mathsf{sgn}(w_j)$$

for $j = s_n + 1, \ldots, d_n$, where

$$\mathcal{L}_n(\mathbf{w}_n) = \big[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_n)\big]^\intercal \big[\mathcal{Y}_n - \hat{\mathcal{M}}(\mathbf{w}_n)\big]$$

and

$$\begin{aligned}
\frac{\partial \mathcal{L}_n(\mathbf{w}_n)}{\partial w_j} &= \mathcal{Y}_n^\intercal \mathcal{S}_n(j)\big[\mathcal{Y}_n - \mathcal{S}_n(\mathcal{Y})\mathbf{w}_n\big] \\
&= \mathcal{Y}_n^\intercal \mathcal{S}_n(j)\big[\mathcal{Y}_n - \mathcal{S}_n(\mathcal{Y})\mathbf{w}_o\big] - \mathcal{Y}_n^\intercal \mathcal{S}_n(j)\mathcal{S}_n(\mathcal{Y})\big(\mathbf{w}_n - \mathbf{w}_o\big) \\
&=: \Xi_{n3} + \Xi_{n4}.
\end{aligned}$$

As in the proof of Theorem 2(i), it is easy to prove that

$$|\Xi_{n3}| = O_P\big(\sqrt{nq_n}\big) \text{ and } |\Xi_{n4}| = O_P\big(\sqrt{nq_n}\big), \tag{B.20}$$

which indicate that

$$\frac{\partial \mathcal{L}_n(\mathbf{w}_n)}{\partial w_j} = O_P\big(\sqrt{nq_n}\big). \tag{B.21}$$

Hence, by (B.21), we have

$$\begin{aligned}
\frac{\partial \mathcal{Q}_n(\mathbf{w}_n)}{\partial w_j} &= O_P\big(\sqrt{q_n n}\big) + np_\lambda'(|w_j|)\mathsf{sgn}(w_j) \\
&= O_P\big(\sqrt{q_n n}\big) + n\lambda\big[\lambda^{-1}p_\lambda'(|w_j|)\mathsf{sgn}(w_j)\big] \\
&= O_P\Big(\sqrt{q_n n}\big\{1 + \frac{\sqrt{n}\lambda}{\sqrt{q_n}}\big[\lambda^{-1}p_\lambda'(|w_j|)\mathsf{sgn}(w_j)\big]\big\}\Big).
\end{aligned} \tag{B.22}$$

Since $\frac{\sqrt{n}\lambda}{\sqrt{q_n}} \to \infty$, we can show that (B.18) and (B.19) hold by using (B.22). We thus complete the proof of Theorem 2(ii).

(iii) Let $\hat{\mathbf{w}}_0^\intercal(n) = (\hat{\mathbf{w}}_n^\intercal(1), \mathbf{0}^\intercal)$ and $\hat{w}_j$ be the estimator of w_{oj} for $j = 1, 2, \ldots, q_n$. By Theorem 2(ii), we have

$$\frac{\partial \mathcal{Q}_n\big(\hat{\mathbf{w}}_n\big)}{\partial w_j} = \frac{\partial \mathcal{Q}_n\big(\hat{\mathbf{w}}_0(n)\big)}{\partial w_j} = 0 \tag{B.23}$$

for $j = 1, 2, \ldots, s_n$. By Taylor's expansion and Theorem 2(ii), we have for $j = 1, 2, \ldots, s_n$,

$$\begin{aligned} \frac{\partial \mathcal{Q}_n\big(\hat{\mathbf{w}}_0(n)\big)}{\partial w_j} &= \frac{\partial \mathcal{Q}_n\big(\mathbf{w}_o\big)}{\partial w_j} + \sum_{l=1}^{q_n} \frac{\partial^2 \mathcal{Q}_n\big(\mathbf{w}_n^*\big)}{\partial w_j \partial w_l}(\hat{w}_l - w_{ol}) \\ &= \frac{\partial \mathcal{Q}_n\big(\mathbf{w}_o\big)}{\partial w_j} + \sum_{l=1}^{s_n} \frac{\partial^2 \mathcal{Q}_n\big(\mathbf{w}_n^*\big)}{\partial w_j \partial w_l}(\hat{w}_l - w_{ol}), \end{aligned} \tag{B.24}$$

where $\mathbf{w}_n^*$ lies between $\hat{\mathbf{w}}_0(n)$ and $\mathbf{w}_o$.

Define

$$\boldsymbol{\Theta}_n^\intercal\big(\mathbf{w}_o\big) = \left[\frac{\partial \mathcal{Q}_n(\mathbf{w}_o)}{\partial w_1}, \ldots, \frac{\partial \mathcal{Q}_n(\mathbf{w}_o)}{\partial w_{s_n}}\right]$$

and $\boldsymbol{\Phi}_n\big(\mathbf{w}_n^*\big)$ be the $s_n \times s_n$ matrix whose (j,k)-th component is $\frac{\partial^2 \mathcal{Q}_n(\mathbf{w}_n^*)}{\partial w_j \partial w_k}$. Then, by (B.24), we have

$$\hat{\mathbf{w}}_n(1) - \mathbf{w}_o(1) = \boldsymbol{\Phi}_n^{-1}\big(\mathbf{w}_n^*\big)\boldsymbol{\Theta}_n\big(\mathbf{w}_o\big). \tag{B.25}$$

Following the proof of Theorem 3.3 in Li *et al* (2015), we can show that

$$\frac{1}{n}\boldsymbol{\Theta}_n\big(\mathbf{w}_o\big) \overset{P}{\sim} \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\xi}_t + \boldsymbol{\omega}_n, \tag{B.26}$$

where $\boldsymbol{\xi}_t$ and $\boldsymbol{\omega}_n$ are defined in Section 3.1. On the other hand, we may also show that

$$\frac{1}{n}\boldsymbol{\Phi}_n\big(\mathbf{w}_n^*\big) \overset{P}{\sim} \boldsymbol{\Lambda}_{n1} + \boldsymbol{\Omega}_n, \tag{B.27}$$

where $\boldsymbol{\Lambda}_{n1}$ and $\boldsymbol{\Omega}_n$ are defined in Section 3.1. Letting $\boldsymbol{u}_{nt} = \mathbf{A}_n\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{\xi}_t$, by (B.25)–(B.27), it suffices to show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{u}_{nt} \xrightarrow{d} \mathsf{N}\big(\mathbf{0}, \mathbf{A}_0\big), \tag{B.28}$$

which can be proved by using the central limit theorem for the stationary α-mixing sequence, for example, Theorem 3.2.1 in Lin and Lu (1996). The proof of Theorem 2(iii) has thus been completed. ■

Proof of Theorem 3 (i) The proof is similar to the proof of Theorem 1 in Bai and Ng (2002) and the proof of Theorem 3.3 in Fan *et al* (2013). By the definition of $\hat{\mathbf{f}}_t$, we readily have

$$\hat{\mathbf{V}}\left(\hat{\mathbf{f}}_t-\mathbf{H}\mathbf{f}_t^0\right) = \frac{1}{np_n}\left(\sum_{s=1}^{n}\sum_{k=1}^{p_n}\hat{\mathbf{f}}_s(\mathbf{f}_s^0)^{\intercal}\mathbf{b}_k^0 u_{tk}+\sum_{s=1}^{n}\sum_{k=1}^{p_n}\hat{\mathbf{f}}_s(\mathbf{f}_t^0)^{\intercal}\mathbf{b}_k^0 u_{sk}+\right.$$
$$\left.\sum_{s=1}^{n}\sum_{k=1}^{p_n}\hat{\mathbf{f}}_s\mathsf{E}\left[u_{sk}u_{tk}\right]+\sum_{s=1}^{n}\sum_{k=1}^{p_n}\hat{\mathbf{f}}_s\left\{u_{sk}u_{tk}-\mathsf{E}\left[u_{sk}u_{tk}\right]\right\}\right) \tag{B.29}$$

for any $1\le t\le n$.

By the conditions B1 and B4, and following the proof of (A.35) in Appendix A, we may show that uniformly for $1\le t\le n$,

$$\frac{1}{np_n}\left(\sum_{s=1}^{n}\sum_{k=1}^{p_n}\hat{\mathbf{f}}_s\mathsf{E}\left[u_{sk}u_{tk}\right]\right)=O_P\left(n^{-1/2}\right) \tag{B.30}$$

and

$$\frac{1}{np_n}\left(\sum_{s=1}^{n}\sum_{k=1}^{p_n}\hat{\mathbf{f}}_s\left\{u_{sk}u_{tk}-\mathsf{E}\left[u_{sk}u_{tk}\right]\right\}\right)=O_P\left(n^{1/4}p_n^{-1/2}\right). \tag{B.31}$$

Noting that $\left\|\sum_{s=1}^{n}\hat{\mathbf{f}}_s(\mathbf{f}_s^0)^{\intercal}\right\|=O_P(n)$ by the condition B2, and $\max_t\|\sum_{k=1}^{p_n}\mathbf{b}_k^0u_{tk}\|=O_P(n^{1/4}p_n^{1/2})$ by (3.9) in the condition B4, we have

$$\frac{1}{np_n}\left(\sum_{s=1}^{n}\sum_{k=1}^{p_n}\hat{\mathbf{f}}_s(\mathbf{f}_s^0)^{\intercal}\mathbf{b}_k^0u_{tk}\right)=O_P\left(n^{1/4}p_n^{-1/2}\right) \tag{B.32}$$

uniformly for $1\le t\le n$.

Notice that

$$\begin{aligned}\max_t\left\|\sum_{s=1}^{n}\sum_{k=1}^{p_n}\hat{\mathbf{f}}_s(\mathbf{f}_t^0)^{\intercal}\mathbf{b}_k^0u_{sk}\right\| &\le \max_t\|\mathbf{f}_t^0\|\left(\sum_{s=1}^{n}\left\|\hat{\mathbf{f}}_s\right\|^2\right)^{1/2}\left(\sum_{s=1}^{n}\left\|\sum_{k=1}^{p_n}\mathbf{b}_k^0u_{sk}\right\|^2\right)^{1/2}\\ &= O_P(1)\cdot O_P\left(n^{1/2}\right)\cdot O_P\left(n^{3/4}p_n^{1/2}\right),\end{aligned}$$

by using the conditions B2 and B4. Hence, we have

$$\frac{1}{np_n}\left(\sum_{s=1}^{n}\sum_{k=1}^{p_n}\hat{\mathbf{f}}_s(\mathbf{f}_t^0)^{\intercal}\mathbf{b}_k^0u_{sk}\right)=O_P\left(n^{1/4}p_n^{-1/2}\right) \tag{B.33}$$

uniformly for $1\le t\le n$.

By the definition of $\hat{\mathcal{F}}_n$ and following the arguments in the proof of Lemma 5 in Appendix A, we may show that

$$\hat{\mathbf{V}} - \left(\frac{1}{n}\hat{\mathcal{F}}_n^{\intercal}\mathcal{F}_n^0\right)\left(\frac{1}{p_n}(\mathbf{B}_n^0)^{\intercal}\mathbf{B}_n^0\right)\left(\frac{1}{n}(\mathcal{F}_n^0)^{\intercal}\hat{\mathcal{F}}_n\right) = o_P(1). \tag{B.34}$$

Furthermore, by Lemma 5 again,

$$\frac{1}{n}(\mathcal{F}_n^0)^{\intercal}(\mathcal{F}_n^0) - \left(\frac{1}{n}(\mathcal{F}_n^0)^{\intercal}\hat{\mathcal{F}}_n\right)\left(\frac{1}{n}\hat{\mathcal{F}}_n^{\intercal}\mathcal{F}_n^0\right) = o_P(1),$$

which together with the condition B2, implies that $\hat{\mathcal{F}}_n^{\intercal}\mathcal{F}_n^0/n$ is asymptotically invertible. By (B.34) and noting that $(\mathbf{B}_n^0)^{\intercal}\mathbf{B}_n^0/p_n$ is positive definite, we may show that $\hat{\mathbf{V}}$ is also asymptotically invertible. We can then complete the proof of (3.12) in Theorem 3(i) by using this fact and (B.29)–(B.33).

(ii) Let

$$\eta_{tk,f}^* = Y_t - m_{k,f}^*(\tilde{f}_{tk}^0) = Y_t - \mathsf{E}\big[Y_t|\tilde{f}_{tk}^0\big],$$

where $\tilde{f}_{tk}^0 = e_r(k)\mathbf{H}\mathbf{f}_t^0$ is defined as in Section 2.2. By the condition B1, given the $r \times r$ matrix $\mathbf{H}$ and for $k = 1, \cdots, r$, $\big\{(\eta_{tk,f}^*, \tilde{f}_{tk}^0), t = 1, \cdots, n\big\}$ is stationary α-mixing. Note that

$$\begin{aligned}
\hat{m}_{k,f}^*(z_k) - \tilde{m}_{k,f}^*(z_k) &= \big[\hat{m}_{k,f}^*(z_k) - m_{k,f}^*(z_k)\big] - \big[\tilde{m}_{k,f}^*(z_k) - m_{k,f}^*(z_k)\big] \\
&= \frac{\sum_{t=1}^n K\big(\frac{\hat{f}_{tk}-z_k}{h_3}\big)\,\big[Y_t - m_{k,f}^*(z_k)\big]}{\sum_{t=1}^n K\big(\frac{\hat{f}_{tk}-z_k}{h_3}\big)} - \frac{\sum_{t=1}^n K\big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\big)\,\big[Y_t - m_{k,f}^*(z_k)\big]}{\sum_{t=1}^n K\big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\big)} \\
&= \left[\frac{\sum_{t=1}^n K\big(\frac{\hat{f}_{tk}-z_k}{h_3}\big)\eta_{tk,f}^*}{\sum_{t=1}^n K\big(\frac{\hat{f}_{tk}-z_k}{h_3}\big)} - \frac{\sum_{t=1}^n K\big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\big)\eta_{tk,f}^*}{\sum_{t=1}^n K\big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\big)}\right] + \\
&\quad \left[\frac{\sum_{t=1}^n K\big(\frac{\hat{f}_{tk}-z_k}{h_3}\big)\Delta_{tk,m}}{\sum_{t=1}^n K\big(\frac{\hat{f}_{tk}-z_k}{h_3}\big)} - \frac{\sum_{t=1}^n K\big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\big)\Delta_{tk,m}}{\sum_{t=1}^n K\big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\big)}\right] \\
&=: \Gamma_{n1}(z_k) + \Gamma_{n2}(z_k),
\end{aligned} \tag{B.35}$$

where $\Delta_{tk,m} = m_{k,f}^*(\tilde{f}_{tk}^0) - m_{k,f}^*(z_k)$.

We first consider the uniform convergence for $\Gamma_{n1}(z_k)$. It is easy to show that

$$\begin{aligned}
\Gamma_{n1}(z_k) &= \left[\frac{\sum_{t=1}^n K\big(\frac{\hat{f}_{tk}-z_k}{h_3}\big)\eta_{tk,f}^*}{\sum_{t=1}^n K\big(\frac{\hat{f}_{tk}-z_k}{h_3}\big)} - \frac{\sum_{t=1}^n K\big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\big)\eta_{tk,f}^*}{\sum_{t=1}^n K\big(\frac{\hat{f}_{tk}-z_k}{h_3}\big)}\right] + \\
&\quad \left[\frac{\sum_{t=1}^n K\big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\big)\eta_{tk,f}^*}{\sum_{t=1}^n K\big(\frac{\hat{f}_{tk}-z_k}{h_3}\big)} - \frac{\sum_{t=1}^n K\big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\big)\eta_{tk,f}^*}{\sum_{t=1}^n K\big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\big)}\right] \\
&=: \Gamma_{n1,1}(z_k) + \Gamma_{n1,2}(z_k).
\end{aligned} \tag{B.36}$$

Following the arguments in the proof of Lemma 3, we may prove

$$\max_{1\leq k\leq r}\sup_{z_k\in\mathcal{F}_k^*}\left|\frac{1}{nh_3}\sum_{t=1}^{n}K\Big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\Big)\eta_{tk,f}^*\right|=O_P\left(\sqrt{\log n/(nh_3)}\right). \tag{B.37}$$

By the condition B5(i), we apply Taylor's expansion to the kernel function:

$$K\Big(\frac{\hat{f}_{tk}-z_k}{h_3}\Big)-K\Big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\Big)=\frac{\hat{f}_{tk}-\tilde{f}_{tk}^0}{h_3}K'\Big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\Big)+\frac{1}{2}\left(\frac{\hat{f}_{tk}-\tilde{f}_{tk}^0}{h_3}\right)^2K''\Big(\frac{\tilde{f}_{tk}^\diamond-z_k}{h_3}\Big),$$

where $\tilde{f}_{tk}^\diamond$ lies between $\tilde{f}_{tk}^0$ and $\hat{f}_{tk}$, $K'(\cdot)$ and $K''(\cdot)$ are the first and second derivatives of the kernel function $K(\cdot)$. By the above Taylor's expansion and Theorem 3(i), we have

$$\begin{aligned}&\max_{1\leq k\leq r}\sup_{z_k\in\mathcal{F}_k^*}\left|\frac{1}{nh_3}\sum_{t=1}^{n}K\Big(\frac{\hat{f}_{tk}-z_k}{h_3}\Big)-\frac{1}{nh_3}\sum_{t=1}^{n}K\Big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\Big)\right|\\ =\;& O_P\left(n^{-1/2}h_3^{-1}+n^{1/4}p_n^{-1/2}h_3^{-1}\right)+O_P\left(n^{-1}h_3^{-3}+n^{1/2}p_n^{-1}h_3^{-3}\right).\end{aligned} \tag{B.38}$$

By (B.37), (B.38), and $n^{1-\gamma_0}h_3^3\to\infty$ and $n=o(p_n^2h_3^{12})$ in the condition B5(ii), we readily have

$$\begin{aligned}\max_{1\leq k\leq r}\sup_{z_k\in\mathcal{F}_k^*}|\Gamma_{n1,2}(z_k)| &= O_P\left(n^{-1}h_3^{-3/2}(\log n)^{1/2}+n^{-1/4}h_3^{-3/2}p_n^{-1/2}(\log n)^{1/2}\right)+\\ &\quad O_P\left(n^{-3/2}h_3^{-7/2}(\log n)^{1/2}+h_3^{-7/2}p_n^{-1}(\log n)^{1/2}\right)\\ &= o_P\left(n^{-1/2}\right).\end{aligned} \tag{B.39}$$

Once again, by the condition B5(i), we apply Taylor's expansion to the kernel function:

$$K\Big(\frac{\hat{f}_{tk}-z_k}{h_3}\Big)-K\Big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\Big)=\frac{\hat{f}_{tk}-\tilde{f}_{tk}^0}{h_3}K'\Big(\frac{\tilde{f}_{tk}^*-z_k}{h_3}\Big), \tag{B.40}$$

where $\tilde{f}_{tk}^*$ lies between $\tilde{f}_{tk}^0$ and $\hat{f}_{tk}$. As in (A.9), by the condition A5 and B5(iii), we may show that $\max_{1\leq k\leq r}\mathsf{E}[\exp\{\varsigma|\eta_{tk,f}^*|\}]<\infty$, which indicates that

$$\max_{1\leq t\leq n}\max_{1\leq k\leq r}|\eta_{tk,f}^*|=o_P(n^{\iota}), \tag{B.41}$$

where $\iota>0$ can be arbitrarily small. Using (B.29), (B.31)–(B.33), (B.40) and (B.41), and noting that the matrix $\hat{\mathbf{V}}$ is asymptotically invertible, we have

$$\begin{aligned}
& \max_{1\leq k\leq r}\sup_{z_k\in\mathcal{F}_k^*}\left|\frac{1}{nh_3}\sum_{t=1}^n \eta_{tk,f}^* K\Big(\frac{\hat{f}_{tk}-z_k}{h_3}\Big) - \frac{1}{nh_3}\sum_{t=1}^n \eta_{tk,f}^* K\Big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\Big)\right| \\
\leq\ & c_\star\cdot\max_{1\leq k\leq r}\sup_{z_k\in\mathcal{F}_k^*}\left|\frac{1}{n^2h_3^2p_n}\sum_{s=1}^n \hat{f}_{sk}\sum_{t=1}^n\sum_{k=1}^{p_n}\mathsf{E}\left[u_{sk}u_{tk}\right]\eta_{tk,f}^* K'\Big(\frac{\tilde{f}_{tk}^*-z_k}{h_3}\Big)\right| + \\
& O_P\left(n^{1/4}p_n^{-1/2}h_3^{-1}\right)\cdot\max_{1\leq k\leq r}\sup_{z_k\in\mathcal{F}_k^*}\frac{1}{nh_3}\sum_{t=1}^n\left|\eta_{tk,f}^*\right|\left|K'\Big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\Big)\right| + \\
& O_P\left(n^{1/4+\iota}p_n^{-1/2}h_3^{-3}(n^{-1/2}+n^{1/4}p_n^{-1/2})\right) \\
=\ & c_\star\cdot\max_{1\leq k\leq r}\sup_{z_k\in\mathcal{F}_k^*}\left|\frac{1}{n^2h_3^2p_n}\sum_{s=1}^n \hat{f}_{sk}\sum_{t=1}^n\sum_{k=1}^{p_n}\mathsf{E}\left[u_{sk}u_{tk}\right]\eta_{tk,f}^* K'\Big(\frac{\tilde{f}_{tk}^*-z_k}{h_3}\Big)\right| + \\
& O_P\left(n^{1/4+\iota}p_n^{-1/2}h_3^{-1}+n^{-1/4+\iota}p_n^{-1/2}h_3^{-3}+n^{1/2+\iota}p_n^{-1}h_3^{-3}\right),
\end{aligned}$$

where $c_\star$ is a positive constant. Letting ι be sufficiently close to zero, as $nh_3^4=O(1)$, $p_nh_3^9\to\infty$, and $n=o(p_n^2h_3^{13})$ in the condition B5(ii), we may show that

$$\begin{aligned}
& \max_{1\leq k\leq r}\sup_{z_k\in\mathcal{F}_k^*}\left|\frac{1}{nh_3}\sum_{t=1}^n \eta_{tk,f}^* K\Big(\frac{\hat{f}_{tk}-z_k}{h_3}\Big) - \frac{1}{nh_3}\sum_{t=1}^n \eta_{tk,f}^* K\Big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\Big)\right| \\
\leq\ & c_\star\cdot\max_{1\leq k\leq r}\sup_{z_k\in\mathcal{F}_k^*}\left|\frac{1}{n^2h_3^2p_n}\sum_{s=1}^n \hat{f}_{sk}\sum_{t=1}^n\sum_{k=1}^{p_n}\mathsf{E}\left[u_{sk}u_{tk}\right]\eta_{tk,f}^* K'\Big(\frac{\tilde{f}_{tk}^*-z_k}{h_3}\Big)\right| + o_P\left(n^{-1/2}\right).
\end{aligned}\tag{B.42}$$

By the conditions B1 and B4, for $\kappa_n=[n^{\gamma_0-\iota}]$ with γ_0 specified in the condition B5(ii), there exists $0<\theta_\star<1$ such that

$$\sum_{t:|s-t|>\kappa_n}\mathsf{E}\left[u_{sk}u_{tk}\right]=O\left(\theta_\star^{7\kappa_n/8}\right),$$

which implies that

$$\max_{1\leq k\leq r}\sup_{z_k\in\mathcal{F}_k^*}\left|\frac{1}{n^2h_3^2p_n}\sum_{s=1}^n \hat{f}_{sk}\sum_{|t-s|>\kappa_n}\sum_{k=1}^{p_n}\mathsf{E}\left[u_{sk}u_{tk}\right]\eta_{tk,f}^* K'\Big(\frac{\tilde{f}_{tk}^*-z_k}{h_3}\Big)\right| = o_P\left(n^{-1/2}\right). \tag{B.43}$$

Using (B.41), (B.43) and Theorem 3(i), and applying Taylor's expansion to $K'(\cdot)$, we have

$$\begin{aligned}
& \max_{1\le k\le r}\sup_{z_k\in\mathcal{F}_k^*}\left|\frac{1}{n^2h_3^2p_n}\sum_{s=1}^n\hat{f}_{sk}\sum_{t=1}^n\sum_{k=1}^{p_n}\mathsf{E}\left[u_{sk}u_{tk}\right]\eta_{tk,f}^*K'\Big(\frac{\tilde{f}_{tk}^*-z_k}{h_3}\Big)\right| \\
= & \max_{1\le k\le r}\sup_{z_k\in\mathcal{F}_k^*}\left|\frac{1}{n^2h_3^2p_n}\sum_{s=1}^n\hat{f}_{sk}\sum_{|t-s|\le\kappa_n}\sum_{k=1}^{p_n}\mathsf{E}\left[u_{sk}u_{tk}\right]\eta_{tk,f}^*K'\Big(\frac{\tilde{f}_{tk}^*-z_k}{h_3}\Big)\right|+o_P\left(n^{-1/2}\right) \\
\le & \max_{1\le k\le r}\sup_{z_k\in\mathcal{F}_k^*}\left|\frac{1}{n^2h_3^2p_n}\sum_{s=1}^n\hat{f}_{sk}\sum_{|t-s|\le\kappa_n}\sum_{k=1}^{p_n}\mathsf{E}\left[u_{sk}u_{tk}\right]\eta_{tk,f}^*K'\Big(\frac{\tilde{f}_{tk}^0-z_k}{h_3}\Big)\right|+ \\
& O_P\left(\frac{\kappa_n}{n^{3/2-\iota}h_3^3}+\frac{\kappa_n}{n^{3/4-\iota}p_n^{1/2}h_3^3}\right)+o_P\left(n^{-1/2}\right) \\
= & O_P\left(\frac{\kappa_n}{n^{1-\iota}h_3}\right)+O_P\left(\frac{\kappa_n}{n^{3/2-\iota}h_3^3}+\frac{\kappa_n}{n^{3/4-\iota}p_n^{1/2}h_3^3}\right)+o_P\left(n^{-1/2}\right)=o_P\left(n^{-1/2}\right),
\end{aligned}\tag{B.44}$$

where we have used the condition B5(ii). By (B.38), (B.42) and (B.44), we may prove that

$$\max_{1\le k\le r}\sup_{z_k\in\mathcal{F}_k^*}\left|\Gamma_{n1,1}(z_k)\right|=o_P\left(n^{-1/2}\right).\tag{B.45}$$

By (B.36), (B.39) and (B.45), we readily have

$$\max_{1\le k\le r}\sup_{z_k\in\mathcal{F}_k^*}\left|\Gamma_{n1}(z_k)\right|=o_P\left(n^{-1/2}\right).\tag{B.46}$$

On the other hand, using Taylor's expansion for $m_{k,f}^*(\cdot)$ and following the arguments in the proofs of (B.39) and (B.45), we may also show that

$$\max_{1\le k\le r}\sup_{z_k\in\mathcal{F}_k^*}\left|\Gamma_{n2}(z_k)\right|=o_P\left(n^{-1/2}\right),\tag{B.47}$$

which, together with (B.46), completes the proof of Theorem 3(ii). ■

Appendix C: An iterative KSIS+PMAMAR procedure

As discussed in Section 4.1 of the main article, the KSIS+PMAMAR procedure can have high false positive rates and low true positive rates when the potential covariates are highly correlated with each other. Hence, we here propose an iterative KSIS+PMAMAR procedure to improve the performance of the non-iterative KSIS+PMAMAR.

Step 1: For each $j = 1, 2, \ldots, p_n + d_n$, estimate the marginal regression function $m_j(x_j)$ by the kernel method and denote the estimate as $\hat{m}_j(x_j)$. Then calculate the sample covariance between Y_t and $\hat{m}_j(X_{tj})$:

$$\hat{\mathsf{v}}(j) = \frac{1}{n}\sum_{t=1}^{n} \hat{m}_j^2(X_{tj}) - \Big[\frac{1}{n}\sum_{t=1}^{n} \hat{m}_j(X_{tj})\Big]^2.$$

Select the variable with the largest $\hat{\mathsf{v}}(j)$ and let

$$\mathcal{S} = \Big\{ j : \ \hat{\mathsf{v}}(j) = \max_i(\hat{\mathsf{v}}(i)), 1 \le i \le p_n + d_n \Big\}.$$

Step 2: Run a linear regression of the response variable Y on the estimated marginal regression functions of the selected variables in $\mathcal{S}$, and obtain the residuals $\widehat{e}^{\mathcal{S}}$.

Step 3: Run a linear regression of the estimated marginal regression function of each variable in $\mathcal{S}^c$, which is defined as $\{1, 2, \ldots, p_n + d_n\} \setminus \mathcal{S}$, on the estimated marginal regression functions of the selected variables in $\mathcal{S}$, and obtain the residuals $\widehat{e}^{i\mathcal{S}}$ for each $i \in \mathcal{S}^c$.

Step 4: Compute the kernel estimate of the marginal regression function, $\widehat{m}_i^e$, of the residuals $\widehat{e}^{\mathcal{S}}$ from Step 2 on the residuals $\widehat{e}^{i\mathcal{S}}$ from Step 3 for each $i \in \mathcal{S}^c$, and calculate the sample covariance $\widehat{\mathsf{v}}^e(i)$ between $\widehat{e}^{\mathcal{S}}$ and $\widehat{m}_i^e$. Add the variable j with the largest $\widehat{\mathsf{v}}^e(i)$ among all $i \in \mathcal{S}^c$ to the set $\mathcal{S}$.

Step 5: Run a PMAMAR regression with the SCAD penalty of Y against X_j, $j \in \mathcal{S}$, as in (2.8), and discard any variables from $\mathcal{S}$ if their corresponding estimated weights are zero.

Step 6: Repeat Steps 2–5 until no new variable is recruited or until the number of variables selected in $\mathcal{S}$ hits a pre-determined number.

In Step 4, we treat the residuals, from the linear regression of the response variable on the marginal regression functions of the variables currently selected, as the new response variable, and the residuals, from linear regression of the marginal regression functions of the unselected variables on those of the selected variables, as the new covariates. We then carry out a nonparametric screening and select the variable with the largest resulting sample covariance $\widehat{\mathsf{v}}^e(i)$ as the candidate to be added to $\mathcal{S}$. The use of the residuals, instead of the original Y and unselected $\widehat{m}_j$'s, reduces the priority of the remaining irrelevant variables, which are highly correlated with some selected relevant variables, being picked, and increases the priority of the remaining relevant variables, which are marginally insignificant but jointly significant, being picked. Hence, this iterative procedure may help reduce false positive rates and increase true positive rates. The variables in the selected set $\mathcal{S}$ then undergo the PMAMAR regression with the SCAD penalty. The set $\mathcal{S}$ is updated by discarding any variables having insignificant weights. Other penalty functions such as the LASSO and the MCP can equally apply in Step 5. The above iterative procedure can be seen as a greedy selection algorithm, since at most one variable is selected in each iteration. It starts with zero variable and keeps adding or

Table D.1: Average and median computation times in seconds on Example 5.1

Example 5.1 Setting	Time	IKSIS+PMAMAR	KSIS+PMAMAR	penGAM	ISIS
$\mathrm{cov}(\mathbf{Z}) = I_{p_n}$	Mean	2.4556	0.1880	13.1994	0.2612
$(p_n, d_n) = (30, 10)$	Median	2.1850	0.1900	12.5600	0.2400
$\mathrm{cov}(\mathbf{Z}) = I_{p_n}$	Mean	10.0434	0.5266	30.6222	0.4518
$(p_n, d_n) = (150, 50)$	Median	11.2150	0.5200	30.4650	0.3900
$\mathrm{cov}(\mathbf{Z}) = C_{\mathbf{Z}}$	Mean	1.5214	0.1930	21.9628	0.2928
$(p_n, d_n) = (30, 10)$	Median	1.0600	0.1900	21.7500	0.2650
$\mathrm{cov}(\mathbf{Z}) = C_{\mathbf{Z}}$	Mean	7.7534	0.5486	40.6612	0.6574
$(p_n, d_n) = (150, 50)$	Median	6.4100	0.5400	38.7150	0.6100

deleting variables until none of the remaining variables are considered significant in the sense of significance of the weights in PMAMAR.

Appendix D: Additional numerical results

D.1 Computational times

We present here the computation times (i.e., times taken by CPU to process the computation) of the various methods considered in Examples 5.1 and 5.2 of the main document. More specifically, we record the average and median time (in seconds) over 50 replications for a single running of each of the methods considered. For the ISIS and the methods involving PMAMAR, the SCAD penalty is used with the concavity tuning parameter set to its default value of 3.7 and the regularisation parameter λ chosen by a ten-fold cross validation. For the penGAM, the number of interior knots used for the B-spline parameterisation is $\lfloor n^{1/5} \rfloor$ and the tuning parameter λ_1 is selected using cross-validation over the grid 2^{s_i}, $s_i = -0.1i$, $i = 0, 1, 2, \ldots, 100$. The SCAD penalised regression is implemented using the R package "ncvreg", the ISIS method implemented using the "SIS" package, the penGAM method implemented using the "penGAM" package. All the computation was carried out on a Windows 7 PC with 64-bit operating system, 2.30GHz Intel Core i5-2500T CPU and 4.0GB RAM.

Tables D.1 and D.2 show that, in Example 5.1, the penGAM is the most time-consuming method, followed by IKSIS+PMAMAR, then ISIS (except when the exogenous covariates are uncorrelated and $(p_n, d_n) = (150, 50)$) and KSIS+PMAMAR. The IKSIS+PMAMAR requires 10-20 times as much time as that for KSIS+PMAMAR. In Example 5.2, the PCA+KSIS+PMAMAR method is the fastest method, closely followed by PCA+PMAMAR, then KSIS+PMAMAR and ISIS, and penGAM is again the most time-consuming method. This shows that the insertion of the KSIS step between the PCA and PMAMAR speeds up the following PMAMAR step (as less variables go through to the PMAMAR step), leading to PCA+KSIS+PMAMAR being overall faster than PCA+PMAMAR.

Table D.2: Average and median computation times in seconds on Example 5.2

Example 5.2 Setting	Time	PCA+PMAMAR	PCA+KSIS+PMAMAR	KSIS+PMAMAR	penGAM	ISIS
$(p_n, d_n) = (30, 10)$	Mean	0.1110	0.0960	0.2322	102.5164	0.6182
	Median	0.1100	0.1000	0.2300	100.6300	0.6150
$(p_n, d_n) = (150, 50)$	Mean	0.1090	0.0964	0.5196	118.5414	1.6820
	Median	0.1100	0.1000	0.5150	116.8250	1.5100

D.2 Estimated models in Example 5.3

We present here the estimated models for the empirical example considered in Section 5.2 of the main article. The response variable Y_t below represents $\Delta \log(\text{CPI}_t)$ and a list of what the exogenous variables $\mathbf{Z}$ represent is given at the end of this section.

The estimated models:

from the IKSIS+PMAMAR:

$$\begin{aligned}Y_t = &-0.1022 + 0.5554\hat{m}_2(Z_{2,t}) + 0.5606\hat{m}_4(Z_{4,t}) + 0.6536\hat{m}_7(Z_{7,t}) + 0.6018\hat{m}_{12}(Z_{12,t})\\ &+ 1.0631\hat{m}_{18}(Z_{18,t}) + 0.7109\hat{m}_{30}(Z_{30,t}) + 0.8016\hat{m}_{43}(Z_{43,t}) + 0.4576\hat{m}_{44}(Z_{44,t})\\ &+ 0.3463\hat{m}_{54,t}(Y_{t-1}) + 0.6327\hat{m}_{57}(Y_{t-4}) + e_t;\end{aligned}$$

from the KSIS+PMAMAR:

$$\begin{aligned}Y_t = &-0.0345 + 0.7373\hat{m}_2(Z_{2,t}) + 0.4408\hat{m}_4(Z_{4,t}) + 0.5033\hat{m}_{16}(Z_{16,t})\\ &+ 0.3373\hat{m}_{55,t}(Y_{t-2}) + 0.7049\hat{m}_{57}(Y_{t-4}) + e_t;\end{aligned}$$

from the PCA+PMAMAR:

$$\begin{aligned}Y_t = &-0.1415 + 0.3676\hat{m}_3(f_{3,t}) + 0.5594\hat{m}_4(f_{4,t}) + 0.2758\hat{m}_5(f_{5,t}) + 0.2330\hat{m}_7(f_{7,t})\\ &+ 0.5945\hat{m}_8(f_{8,t}) + 0.7685\hat{m}_{13}(f_{13,t}) + 0.3424\hat{m}_{14}(f_{14,t}) + 0.6223\hat{m}_{15}(f_{15,t})\\ &+ 0.4572\hat{m}_{17}(f_{17,t}) + 0.6469\hat{m}_{18}(f_{18,t}) + 0.5060\hat{m}_{20}(f_{20,t}) + 0.7785\hat{m}_{21}(f_{21,t})\\ &+ 0.2659\hat{m}_{22}(f_{22,t}) + 0.9304\hat{m}_{23}(f_{23,t}) + 0.5415\hat{m}_{24}(Y_{t-1}) + 0.3541\hat{m}_{26}(Y_{t-3})\\ &+ 0.5874\hat{m}_{27}(Y_{t-4}) + e_t,\end{aligned}$$

where $f_{i,t}$ are the common factors extracted from the 53 exogenous covariates which together account for 90% of total variance;

from the penGAM:

$$
\begin{aligned}
Y_t =& \hat{m}_1(Z_{1,t}) + \hat{m}_2(Z_{2,t}) + \hat{m}_3(Z_{3,t}) + \hat{m}_4(Z_{4,t}) + \hat{m}_8(Z_{8,t}) + \hat{m}_9(Z_{9,t}) + \hat{m}_{10}(Z_{10,t}) + \hat{m}_{11}(Z_{11,t}) \\
&+ \hat{m}_{13}(Z_{13,t}) + \hat{m}_{14}(Z_{14,t}) + \hat{m}_{16}(Z_{16,t}) + \hat{m}_{17}(Z_{17,t}) + \hat{m}_{18}(Z_{18,t}) + \hat{m}_{21}(Z_{21,t}) + \hat{m}_{22}(Z_{22,t}) \\
&+ \hat{m}_{23}(Z_{23,t}) + \hat{m}_{26}(Z_{26,t}) + \hat{m}_{29}(Z_{29,t}) + \hat{m}_{30}(Z_{30,t}) + \hat{m}_{31}(Z_{31,t}) + \hat{m}_{32}(Z_{32,t}) + \hat{m}_{37}(Z_{37,t}) \\
&+ \hat{m}_{39}(Z_{39,t}) + \hat{m}_{41}(Z_{41,t}) + \hat{m}_{42}(Z_{42,t}) + \hat{m}_{43}(Z_{43,t}) + \hat{m}_{44}(Z_{44,t}) + \hat{m}_{46}(Z_{46,t}) + \hat{m}_{48}(Z_{48,t}) \\
&+ \hat{m}_{52}(Z_{52,t}) + \hat{m}_{53}(Z_{53,t}) + \hat{m}_{54}(Y_{t-1}) + \hat{m}_{57}(Y_{t-4}) + e_t;
\end{aligned}
$$

from the ISIS:

$$
\begin{aligned}
Y_t =& 0.0149 + 0.1435 Z_{1,t} + 0.7354 Z_{2,t} + 0.0846 Z_{4,t} - 0.0846 Z_{21,t} + 0.0933 Z_{37,t} \\
&+ 0.1133 Z_{39,t} + 0.1144 Z_{42,t} + 0.0624 Z_{44,t} + 0.1199 Z_{46,t} - 0.2417 Z_{52,t} \\
&- 0.1916 Y_{t-1} + 0.0526 Y_{t-4} + e_t;
\end{aligned}
$$

from the Phillips curve:

$$
I_{t+1} - I_t = 0.0453 + \hat{\beta}(L) U_t + \hat{\gamma}(L) \Delta I_t + e_{t+1}
$$

where $\hat{\beta}(L) = -1.6654 + 1.4004L + 1.0231L^2 - 0.8242L^3$ (L is the lag operator) and $\hat{\gamma}(L) = -0.9323 - 0.7264L - 0.5693L^2 - 0.0875L^3$;

from the AR model:

$$
Y_t = -0.0349 Y_{t-1} + 0.0640 Y_{t-2} - 0.0049 Y_{t-3} + 0.2533 Y_{t-4} + e_t;
$$

from the VAR modelling:

$$
\begin{pmatrix} Y_t \\ Z_{13,t} \\ Z_{35,t} \\ Z_{38,t} \\ Z_{48,t} \end{pmatrix} = \begin{pmatrix} -0.0846 & -0.0212 & -0.1930 & 0.2016 & -0.0438 \\ -0.2808 & 0.2010 & 0.1541 & 0.3786 & -0.0837 \\ -0.2035 & 0.1250 & 0.1973 & 0.0376 & -0.3542 \\ -0.2358 & 0.3398 & 0.0688 & 0.3870 & -0.4746 \\ -0.0340 & 0.0785 & 0.3505 & 0.3124 & 0.3143 \end{pmatrix} \begin{pmatrix} Y_{t-1} \\ Z_{13,t-1} \\ Z_{35,t-1} \\ Z_{38,t-1} \\ Z_{48,t-1} \end{pmatrix} + \mathbf{e}_t;
$$

from the STAR modelling (a single regime AR model is selected):

$$
Y_t = 0.0180 - 0.0352 Y_{t-1} + 0.0890 Y_{t-2} + e_t.
$$

The following lists the time series used in the methods above to construct estimates and forecasts of inflation. The following symbols are used to denote types of transformation on the data: L=logarithm;

D=first difference; DL=first difference of logarithm. The abbreviation SA stands for seasonally adjusted and NSA for not seasonally adjusted.

Z_1	Durable goods: total IDEF (SA, 2010=100)^{DL}
Z_2	Non-durable goods: total IDEF (SA, 2010=100)^{DL}
Z_3	Services: total IDEF (SA, 2010=100)^{DL}
Z_4	Semi-durable goods: total IDEF (SA, 2010=100)^{DL}
Z_5	UK workforce jobs (SA, thousands)^{DL}
Z_6	LFS: total actual weekly hours worked (SA, millions)^{DL}
Z_7	LFS: unemployed up to 6 months: aged 16 and over (SA, thousands)^{DL}
Z_8	LFS: unemployed over 6 and up to 12 months: aged 16 and over (SA, thousands)^{DL}
Z_9	LFS: unemployed over 12 months: aged 16 and over (SA, thousands)^{DL}
Z_{10}	LFS: unemployed over 24 months: aged 16 and over (SA, thousands)^{DL}
Z_{11}	Net sector output of manufactured products (SA, 2010=100)^{DL}
Z_{12}	Materials and fuels purchased other than FBTP industries (NSA, 2010=100)^{DL}
Z_{13}	IOP: production (SA, 2011=100)^{DL}
Z_{14}	IOP: mining and quarrying (SA, 2011=100)^{DL}
Z_{15}	IOP: manufacturing (SA, 2011=100)^{DL}
Z_{16}	IOP: manufacture of food products beverages and tobacco (SA, 2011=100)^{DL}
Z_{17}	IOP: manufacture of textiles wearing apparel and leather products (SA, 2011=100)^{DL}
Z_{18}	IOP: manufacture of basic pharmaceutical products and pharmaceutical preparations (SA, 2011=100)^{DL}
Z_{19}	IOP: manufacture of wood and paper products and printing (SA, 2011=100)^{DL}
Z_{20}	IOP: manufacture of electrical equipment (SA, 2011=100)^{DL}
Z_{21}	IOP: manufacture of coke and refined petroleum product (SA, 2011=100)^{DL}
Z_{22}	IOP: manufacture of rubber plastic products and other non-metallic mineral products (SA, 2011=100)^{DL}
Z_{23}	IOP: other manufacturing and repair (SA, 2011=100)^{DL}
Z_{24}	IOP: manufacture of machinery and equipment (SA, 2011=100)^{DL}
Z_{25}	IOP: manufacture of basic metals and metal products (SA, 2011=100)^{DL}
Z_{26}	IOP: manufacture of computer electronics and optical products (SA, 2011=100)^{DL}
Z_{27}	IOP: manufacture of transport equipment (SA, 2011=100)^{DL}
Z_{28}	RSI: predominantly food stores: all business index (SA, 2011=100)^{DL}
Z_{29}	LFS: in employment: all aged 16 and over (SA, thousands)^{DL}
Z_{30}	CBI: new domestic order volume future down (NSA)^{DL}
Z_{31}	CBI: new domestic order volume future same (NSA)^{DL}
Z_{32}	CBI: new domestic order volume future up (NSA)^{DL}

Z_{33}	Exchange rate: Swiss Franc to UK Sterling (NSA)^{DL}
Z_{34}	Exchange rate: US Dollar to UK Sterling (NSA)^{DL}
Z_{35}	UK exchange rate effective: Sterling (NSA)^{DL}
Z_{36}	UK nationwide house price index of all properties (SA)^{DL}
Z_{37}	FTSE all share price index^{DL}
Z_{38}	Crude oil: Brent dated FOB (USD/BBL)^{DL}
Z_{39}	LFS: unemployment rate: all aged 16 and over (SA, %)
Z_{40}	CBI: trades: average selling price reported (NSA)
Z_{41}	CBI: trades: business situation (NSA)
Z_{42}	Changes in inventories including alignment adjustment (SA, million pounds)
Z_{43}	Change in inventories: wholesale (SA, million pounds)
Z_{44}	Change in inventories: manufacturing work in progress (SA, million pounds)
Z_{45}	BCC: manufacturing sales: home orders (NSA)^{D}
Z_{46}	BCC: manufacturing sales: home sales (NSA)^{D}
Z_{47}	CBI: business optimism (NSA)^{D}
Z_{48}	Bank of England base rate (NSA)^{D}
Z_{49}	UK interbank lending rate: 3 month mean LIBID/LIBOR (NSA)^{D}
Z_{50}	CBI: export order book volume balance (NSA)^{D}
Z_{51}	CBI: domestic deliveries: next quarter balance (NSA)^{D}
Z_{52}	CBI: domestic deliveries: past quarter balance (NSA)^{D}
Z_{53}	BCC: manufacturing full capacity (NSA)^{L}

References

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.

Bosq, D., 1998. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer.

Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96, 1348–1360.

Fan, J., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements (with discussions). *Journal of the Royal Statistical Society: Series B* 75, 603–680.

Fan, J., Peng, H., 2004. Nonconcave penalized likelihood with a diverging number of parameters. *Annals of Statistics* 32, 928–961.

Li, D., Linton, O., Lu, Z., 2015. A flexible semiparametric forecasting model for time series. *Journal of Econometrics* 187, 345–357.

Li, D., Lu, Z., Linton, O., 2012. Local linear fitting under near epoch dependence: uniform consistency with convergence rates. *Econometric Theory* 28, 935–958.

Lin, Z., Lu, C., 1996. *Limit Theory for Mixing Dependent Random Variables*. Science Press/Kluwer Academic Publishers.