TIME-VARYING ESTIMATION AND DYNAMIC MODEL SELECTION WITH AN APPLICATION OF NETWORK DATA

Lan Xue, Xinxin Shu and Annie Qu

Oregon State University, Merck and University of Illinois at Urbana-Champaign

Supplementary Material

A. Notations

In the following, let C_κ be the space of κ-times continuously differentiable functions on $[0,1]$, and G_n be the spline approximation space of order κ and knot sequence Υ_n. Let $\|\cdot\|$ be the usual vector or function L_2 norm, unless otherwise defined. In the following, we use $\{\sigma^{ii}(t)\}_{i=1}^p$ and $\{\widehat{\sigma}^{ii}(t)\}_{i=1}^p$ to denote the true function and an initial estimator that satisfies condition (C2), respectively. For notation simplicity, we assume $w_{i_{ku}} = 1$ in the following. The proof for the general form of w_{iu} follows similarly under condition (C1). We denote any positive constants by the same letters c, C without distinction in each case.

Let M be the model space as a collection of vectors of functions each with $p\,(p-1)\,/2$ elements,

$$M = \left\{\rho\,(t) = \left(\rho^{ij}\,(t)\,, 1 \le i < j \le p\right), \rho^{ij}\,(t) \in C_\kappa\right\},$$

and let the approximation space be defined similarly as,

$$M_n = \left\{g\,(t) = \left(g^{ij}\,(t)\,, 1 \le i < j \le p\right), g^{ij}\,(t) \in G_n\right\}.$$

For any $\rho \in M$, we define the theoretical and empirical norms on M respectively as

$$\|\rho\|_2^2 = E\left[\sum_{i=1}^{p}\left(\sum_{j\neq i}\rho^{ij}(T)\sqrt{\frac{\sigma^{jj}(T)}{\sigma^{ii}(T)}}Y_j(T)\right)^2\right],$$

and

$$\|\rho\|_n^2 = \frac{1}{nm}\sum_{k=1}^{n}\sum_{u=1}^{m}\sum_{i=1}^{p}\left(\sum_{j\neq i}^{p}\rho^{ij}(t_{ku})\sqrt{\frac{\sigma^{jj}(t_{ku})}{\sigma^{ii}(t_{ku})}}y_j^k(t_{ku})\right)^2.$$

B. Technical Lemmas

Lemma 1. *Under condition (C3), there exist constants* $C > c > 0$, *such that*

$$C\sum_{1\leq i<j\leq p}\left\|\rho^{ij}(T)\right\|_2^2 \geq \|\rho\|_2^2 \geq c\sum_{1\leq i<j\leq p}\left\|\rho^{ij}(T)\right\|_2^2,$$

where $\|\rho^{ij}(T)\|_2^2 = E\left(\rho^{ij}(T)\right)^2$.

Proof: Let $\rho^i(T) = \left(\rho^{ij}(T), j\neq i\right)$, and $\widetilde{\mathbf{Y}}_i(T) = \left(\sqrt{\frac{\sigma^{jj}(t)}{\sigma^{ii}(t)}}Y_j(T), j\neq i\right)$. Then by condition (C3), there exists a constant $c > 0$, such that

$$\begin{aligned}\|\rho\|_2^2 &= E\left[\sum_{i=1}^{p}\left(\rho^i(T)\right)^T\widetilde{\mathbf{Y}}_i^T(T)\,\widetilde{\mathbf{Y}}_i(T)\,\rho^i(T)\right] \geq cE\left[\sum_{i=1}^{p}\left(\rho^i(T)\right)^T\rho^i(T)\right]\\ &= cE\left[\sum_{i=1}^{p}\sum_{j\neq i}\left(\rho^{ij}(T)\right)^2\right] = 2c\sum_{1\leq i<j\leq p}E\left(\rho^{ij}(T)\right)^2 = 2c\sum_{1\leq i<j\leq p}\left\|\rho^{ij}(T)\right\|_2^2.\end{aligned}$$

The other side of the inequality follows similarly from condition (C3). □

Lemma 2. *Under conditions (C3), (C4), (C5) and (C7), there exist constants* $c, C > 0$ *such that, except on an event whose probability goes to zero, as* $n\to\infty$, *one has,*

$$c\|\rho\|_2^2 \leq \|\rho\|_n^2 \leq C\|\rho\|_2^2.$$

Proof: The proof follows similarly from Lemma A.4 of Xue and Yang (2006). □

Lemma 3. *Given conditions (C1), (C2), (C5), (C6) and (C8), there exist constants $C > c > 0$ such that, except on an event whose probability goes to zero, as $n \to \infty$, for any vector β_n of length $p(p-1)J_n/2$, one has,*

$$\frac{c}{N_n}\|\beta_n\|^2 \leq \frac{1}{nm}\beta_n^T\mathcal{X}_n^T\mathcal{X}_n\beta_n \leq \frac{C}{N_n}\|\beta_n\|^2.$$

Proof: Write $\beta_n = \left(\beta^{ij}, 1 \leq i < j \leq p\right)^T$, and $\beta^{ij} = \left(\beta_h^{ij}, h = 1, \ldots J_n\right)^T$. Let $g^{ij} = \sum_{h=1}^{J_n}\beta_h^{ij}B_h \in G_n$, and $g = \left(g^{ij}(t), 1 \leq i < j \leq p\right) \in M_n$. Then $\|g\|_n^2 = \frac{1}{nm}\beta_n^T\mathcal{X}_n^T\mathcal{X}_n\beta_n$. By Lemmas 1 and 2, one has that there exist constants $C \geq c > 0$, such that

$$c\sum_{1\leq i<j\leq p}\left\|g^{ij}\right\|_2^2 \leq \frac{1}{nm}\beta_n^T\mathcal{X}_n^T\mathcal{X}_n\beta_n \leq C\sum_{1\leq i<j\leq p}\left\|g^{ij}\right\|_2^2,$$

in which $\|g^{ij}\|_2^2 = E\left(\sum_{h=1}^{J_n}\beta_h^{ij}B_h(T)\right)^2$. Furthermore, Theorem 5.4.2 of Devore and Lorentz (1993) entails that there exists a constant $c > 0$, such that

$$E\left(\sum_{h=1}^{J_n}\beta_h^{ij}B_h(T)\right)^2 \geq c\sum_{h=1}^{J_n}\left(\beta_h^{ij}\right)^2 E\left(B_h^2(T)\right) \geq \frac{c}{N_n}\sum_{h=1}^{J_n}\left(\beta_h^{ij}\right)^2.$$

Therefore, $\frac{1}{nm}\beta_n^T\mathcal{X}_n^T\mathcal{X}_n\beta_n \geq \frac{c}{N_n}\|\beta_n\|^2$ for some $c > 0$. The other side of the inequality follows similarly from the Cauchy-Schwarz inequality. □

Let $G_{ij}^{(o)} = \left\{g = \sum_{h=1}^{N_n+q+1}\beta_h B_h \in G_n, \quad \beta_h = 0, \text{ for } h = \nu_{l_1^{ij}}, \ldots, \nu_{l_2^{ij}} + q\right\}$, and $G_{ij}^{(o)} \subset G_n$ is the oracle spline approximation space containing spline functions with zero values on the null region E^{ij}.

Lemma 4. *Under conditions (C5)-(C7), there exists a spline function $g_{ij}^{(o)} \in G_{ij}^{(o)}$, such that* $\sup_{0<t<1}\left|\rho^{ij}(t) - g_{ij}^{(o)}(t)\right| = O\left(N_n^{-1}\right)$.

Proof: The approximation theory in de Boor (2001) entails that there exists a spline function $g_{ij} \in G_n$ such that $\sup_{0<t<1} |\rho^{ij}(t) - g_{ij}(t)| = O\left(N_n^{-(q+1)}\right)$, where $g_{ij} = \sum_{h=1}^{N_n+q+1} \beta_h^{ij} B_h$ for a set of coefficients $\left\{\beta_h^{ij}\right\}_{h=1}^{N_n+q+1}$. Now let $g_{ij}^* = \sum_{h \in J_{ij}} \beta_h^{ij} B_h$. Let $A_{ij} = J_{ij}^c \backslash E_{ij}$.Then $g_{ij}^* \in G_{ij}^{(o)}$, and $\sup_{t \in E_{ij}} \left|\rho^{ij} - g_{ij}^*\right| = 0, \sup_{t \in J_{ij}} \left|\rho^{ij} - g_{ij}^*\right| = O\left(N_n^{-(q+1)}\right)$, and

$$\begin{aligned}
\sup_{t \in A_{ij}} \left|\rho^{ij} - g_{ij}^*\right| &\leq \sup_{t \in A_{ij}} \left|\rho^{ij} - g_{ij}\right| + \sup_{t \in A_{ij}} \left|g^{ij} - g_{ij}^*\right| \\
&\leq 2 \sup_{0<t<1} \left|\rho^{ij} - g_{ij}\right| + \sup_{t \in A_{ij}} \left|\rho^{ij}\right| \\
&= O\left(N_n^{-(q+1)} + N_n^{-1}\right) = O\left(N_n^{-1}\right).
\end{aligned}$$

Putting these three cases together, one has $\sup_{0<t<1} \left|\rho^{ij} - g_{ij}^*\right| = O\left(N_n^{-1}\right)$. $\square$

Let

$$L(\boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{t}, \mathbf{y}) = \frac{1}{2nm} \sum_{k=1}^{n} \sum_{i=1}^{p} \sum_{u=1}^{m} w_{i_{ku}} \left(y_i^k(t_{ku}) - \sum_{j \neq i}^{p} \sum_{h=1}^{J_n} \beta_h^{ij} B_h(t_{ku}) \sqrt{\frac{\sigma^{jj}(t_{ku})}{\sigma^{ii}(t_{ku})}} y_j^k(t_{ku}) \right)^2,$$

and

$$L_{hij}^{'}(\boldsymbol{\beta}, \boldsymbol{\sigma}) = \partial L(\boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{t}, \mathbf{y}) / \partial \beta_h^{ij},$$

and

$$L_{hij,h'i'j'}^{''}(\boldsymbol{\beta}, \boldsymbol{\sigma}) = \partial^2 L(\boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{t}, \mathbf{y}) / \partial \beta_h^{ij} \partial \beta_{h'}^{i'j'}.$$

Lemma 5. *For any $\widehat{\boldsymbol{\sigma}}(t)$ that satisfies condition (C2), one has,*

$$\max_{h,i,j} \left| L_{hij}^{'}(\boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{t}, \mathbf{y}) - L_{hij}^{'}(\boldsymbol{\beta}, \widehat{\boldsymbol{\sigma}}, \mathbf{t}, \mathbf{y}) \right| = O_p\left(\sqrt{\frac{log(nm)}{nmN_n}} \right),$$

$$\max_{h,i,j,h',i',j'} \left| L_{hij,h'i'j'}^{''}(\boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{t}, \mathbf{y}) - L_{hij,h'i'j'}^{''}(\boldsymbol{\beta}, \widehat{\boldsymbol{\sigma}}, \mathbf{t}, \mathbf{y}) \right| = O_p\left(\sqrt{\frac{log(nm)}{nmN_n}} \right).$$

Proof: Note that

$$
\begin{aligned}
& L'_{hij}(\boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{t}, \mathbf{y}) - L'_{hij}(\boldsymbol{\beta}, \widehat{\boldsymbol{\sigma}}, \mathbf{t}, \mathbf{y}) \\
= & -\frac{1}{nm} \sum_{k=1}^{n} \sum_{u=1}^{m} w_{i_{ku}} B_h(t_{ku}) y_j^k(t_{ku}) y_i^k(t_{ku}) \left(\sqrt{\frac{\sigma^{jj}(t_{ku})}{\sigma^{ii}(t_{ku})}} - \sqrt{\frac{\widehat{\sigma}^{jj}(t_{ku})}{\widehat{\sigma}^{ii}(t_{ku})}} \right) \\
& -\frac{1}{nm} \sum_{k=1}^{n} \sum_{u=1}^{m} w_{i_{ku}} B_h(t_{ku}) y_j^k(t_{ku}) \times
\end{aligned}
$$

$$
\begin{aligned}
& \left[\sum_{j' \neq i}^{p} \sum_{h=1}^{J_n} \beta_h^{ij} B_h(t_{ku}) \left(\sqrt{\frac{\sigma^{j'j'}(t_{ku})}{\sigma^{ii}(t_{ku})}} \sqrt{\frac{\sigma^{jj}(t_{ku})}{\sigma^{ii}(t_{ku})}} - \sqrt{\frac{\widehat{\sigma}^{j'j'}(t_{ku})}{\widehat{\sigma}^{ii}(t_{ku})}} \sqrt{\frac{\widehat{\sigma}^{jj}(t_{ku})}{\widehat{\sigma}^{ii}(t_{ku})}} \right) y_j^k(t_{ku}) \right] \\
& -\frac{1}{nm} \sum_{k=1}^{n} \sum_{u=1}^{m} w_{j_{ku}} B_h(t_{ku}) y_j^k(t_{ku}) y_i^k(t_{ku}) \left(\sqrt{\frac{\sigma^{ii}(t_{ku})}{\sigma^{jj}(t_{ku})}} - \sqrt{\frac{\widehat{\sigma}^{ii}(t_{ku})}{\widehat{\sigma}^{jj}(t_{ku})}} \right) \\
& -\frac{1}{nm} \sum_{k=1}^{n} \sum_{u=1}^{m} w_{j_{ku}} B_h(t_{ku}) y_i^k(t_{ku}) \times \\
& \left[\sum_{j' \neq j}^{p} \sum_{h=1}^{J_n} \beta_h^{ij} B_h(t_{ku}) \left(\sqrt{\frac{\sigma^{j'j'}(t_{ku})}{\sigma^{jj}(t_{ku})}} \sqrt{\frac{\sigma^{ii}(t_{ku})}{\sigma^{jj}(t_{ku})}} - \sqrt{\frac{\widehat{\sigma}^{j'j'}(t_{ku})}{\widehat{\sigma}^{jj}(t_{ku})}} \sqrt{\frac{\widehat{\sigma}^{ii}(t_{ku})}{\widehat{\sigma}^{jj}(t_{ku})}} \right) y_i^k(t_{ku}) \right].
\end{aligned}
$$

By conditions (C2) and (C3), for any $1 \leq i \neq j \leq p$, one has,

$$
\sup_t \left| \sqrt{\frac{\sigma^{jj}(t)}{\sigma^{ii}(t)}} - \sqrt{\frac{\widehat{\sigma}^{jj}(t)}{\widehat{\sigma}^{ii}(t)}} \right| \leq c \sqrt{\frac{log(nm) N_n}{nm}},
$$

$$
\sup_{1 \leq j' \neq i \leq p} \sup_t \left| \sqrt{\frac{\sigma^{j'j'}(t_{ku})}{\sigma^{ii}(t_{ku})}} \sqrt{\frac{\sigma^{jj}(t_{ku})}{\sigma^{ii}(t_{ku})}} - \sqrt{\frac{\widehat{\sigma}^{j'j'}(t_{ku})}{\widehat{\sigma}^{ii}(t_{ku})}} \sqrt{\frac{\widehat{\sigma}^{jj}(t_{ku})}{\widehat{\sigma}^{ii}(t_{ku})}} \right| \leq c \sqrt{\frac{log(nm) N_n}{nm}}.
$$

Furthermore, by conditions (C1), (C4) and bounded property of B-spline basis, one has $\max_{h,i,j} \left| \frac{1}{nm} \sum_{k=1}^{n} \sum_{u=1}^{m} w_{i_{ku}} B_h(t_{ku}) y_j^k(t_{ku}) y_i^k(t_{ku}) \right| = O_p(1/N_n)$, and

$$
\max_{h,i,j} \left| \frac{1}{nm} \sum_{k=1}^{n} \sum_{u=1}^{m} w_{i_{ku}} B_h(t_{ku}) y_j^k(t_{ku}) \left[\sum_{j \neq i}^{p} \sum_{h=1}^{J_n} \beta_h^{ij} B_h(t_{ku}) y_j^k(t_{ku}) \right] \right| = O_p(1/N_n).
$$

Therefore,

$$\max_{h,i,j}\left|L^{'}(\boldsymbol{\beta},\boldsymbol{\sigma},\mathbf{t},\mathbf{y})-L^{'}(\boldsymbol{\beta},\widehat{\boldsymbol{\sigma}},\mathbf{t},\mathbf{y})\right|=O_p\left(\sqrt{\frac{log(nm)}{nmN_n}}\right).$$

The proof of the second order derivative follows similarly.

Lemma 6. *Suppose conditions (C1)-(C6) hold. Let* $Z_h^{ijk}(t)=B_h(t)\sqrt{\frac{\widehat{\sigma}^{jj}(t)}{\widehat{\sigma}^{ii}(t)}}y_j^k(t)$, *and for any* $1\leq i<j\leq p$, *let*

$$\widetilde{\mathbf{c}}_h^{ij}(\beta)=-\frac{1}{nm}\sum_{k=1}^{n}\sum_{u=1}^{m}Z_h^{ijk}(t_{ku})\left(y_i^k(t_{ku})-\sum_{j'\neq i}^{p}\sum_{h\in J_{ij}}\beta_h^{ij'}Z_h^{ij'k}(t_{ku})\right),$$

and $\mathbf{c}_h^{ij}(\beta)=\widetilde{\mathbf{c}}_h^{ij}(\beta)+\widetilde{\mathbf{c}}_h^{ji}(\beta)$. *Then for any* η_n *that satisfies* $\left(N_n^{-(q+2)}+\lambda_n/N_n\right)/\eta_n\to 0$ *and* $\frac{1}{\eta_n}\sqrt{\frac{\log(nm)}{N_nnm}}\to 0$, *one has*

$$P\left(\max_{1\leq i<j\leq p,h\in J_{ij}^c}\left|\mathbf{c}_h^{ij}\left(\widehat{\boldsymbol{\beta}}^{(0)}\right)\right|\geq\eta_n\right)\to 0.$$

Proof: By Lemma 5 and conditions on η_n that $\frac{1}{\eta_n}\sqrt{\frac{log(nm)}{nmN_n}}\to 0$, it is sufficient to consider $\widetilde{\mathbf{c}}_h^{ij}(\beta)$ and $\mathbf{c}_h^{ij}(\beta)$ with $\widehat{\boldsymbol{\sigma}}(t)$ replaced by $\boldsymbol{\sigma}(t)$. By Lemma 4, there exists a constant $c>0$ and spline functions $s^{ij}\in G_{ij}^{(o)}$, such that

$$\max_{1\leq i<j\leq p}\sup_{0<t<1}\left|\rho^{ij}(t)-s^{ij}(t)\right|\leq c\left(N_n^{-(q+1)}+\lambda_n\right). \tag{S0.1}$$

Let $e_h^{ij}=-\frac{1}{nm}\sum_{k=1}^{n}\sum_{u=1}^{m}Z_h^{ijk}(t_u)\left(y_i^k(t_u)-\sum_{j'\neq i}^{p}\rho^{ij'}(t_u)\sqrt{\frac{\sigma^{j'j'}(t_u)}{\sigma^{ii}(t_u)}}y_{j'}^k(t_u)\right)$,

$$\delta_h^{ij}=\frac{1}{nm}\sum_{k=1}^{n}\sum_{u=1}^{m}Z_h^{ijk}(t_u)\left(\sum_{j'\neq i}^{p}\left[s^{ij'}(t_u)-\rho^{ij'}(t_u)\right]\sqrt{\frac{\sigma^{j'j'}(t_u)}{\sigma^{ii}(t_u)}}y_{j'}^k(t_u)\right),$$

and $\varepsilon_h^{ij}(\beta)=-\frac{1}{nm}\sum_{k=1}^{n}\sum_{u=1}^{m}Z_h^{ijk}(t_u)\left\{\sum_{j'\neq i}^{p}\left[s^{ij'}(t_u)-\sum_{h\in J_{ij}}\beta_h^{ij'}B_h^{ij'}(t_u)\right]\sqrt{\frac{\sigma^{j'j'}(t_u)}{\sigma^{ii}(t_u)}}y_{j'}^k(t_u)\right\}$.

Then one has

$$\widetilde{\mathbf{c}}_h^{ij}\left(\beta\right) = e_h^{ij} + \delta_h^{ij} + \varepsilon_h^{ij}\left(\beta\right).$$

Following similar arguments of Lemma 7 in Xue and Qu (2012) and Lemma 5, there exists a $c > 0$ such that

$$E\left(\max_{1\leq i<j\leq p, h\in J_{ij}^c}\left|e_h^{ij}\right|\right) \leq c\sqrt{\log\left(N_n\right)/\left(N_n nm\right)}.$$

Therefore, by Markov's inequality, one has

$$P\left(\max_{1\leq i<j\leq p, h\in J_{ij}^c}\left|e_h^{ij}\right| > \frac{\eta_n}{2}\right) \leq \frac{2c}{\eta_n}\sqrt{\frac{\log\left(N_n\right)}{N_n nm}} \to 0, \tag{S0.2}$$

as $n \to \infty$. On the other hand,

$$\begin{aligned}
\max_{1\leq i<j\leq p, h\in J_{ij}^c}\left|\delta_h^{ij}\right| &\leq \max_{1\leq i<j\leq p, h\in J_{ij}^c}\frac{1}{nm}\sum_{k=1}^{n}\sum_{u=1}^{m}Z_h^{ijk}\left(t_u\right)\left|\sum_{l\neq i}^{p}\left[s^{il}(t_u)-\rho^{il}(t_u)\right]\sqrt{\frac{\widehat{\sigma}^{ll}(t_u)}{\widehat{\sigma}^{ii}(t_u)}}y_l^k(t_u)\right| \\
&\leq cN_n^{-1}\max_{1\leq i<j\leq p, h\in J_{ij}^c}\frac{1}{nm}\sum_{k=1}^{n}\sum_{u=1}^{m}\left|Z_h^{ijk}\left(t_u\right)\right|\sum_{l\neq i}^{p}\sqrt{\frac{\widehat{\sigma}^{ll}(t_u)}{\widehat{\sigma}^{ii}(t_u)}}\left|y_l^k(t_u)\right| \\
&\leq cN_n^{-1}/N_n = o_p\left(\eta_n\right).
\end{aligned} \tag{S0.3}$$

Finally, by the definition of $\widehat{\beta}^{(0)}$ and the fact that $s^{ij} \in G_{ij}^{(o)}$, one has $\varepsilon_h^{ij}\left(\widehat{\beta}^{(0)}\right) = o_p\left(\eta_n\right)$, for $h \in J_{ij}^c$. Then Lemma 6 follows from (S0.2) and (S0.3). □

C. Construction of initial estimators for $\left\{\sigma^{ii}\left(t\right)\right\}_{i=1}^p$.

To prove our asymptotic results, we assume there exist initial estimators of $\{\sigma^{ii}\left(t\right)\}_{i=1}^p$ that satisfy condition (C2). The following shows how such initial estimators can be constructed. For each fixed $i = 1, \ldots, p$, note that $y_i\left(t\right) = \sum_{j\neq i}\rho^{ij}\left(t\right)\sqrt{\left(\sigma^{jj}\left(t\right)/\sigma^{ii}\left(t\right)\right)}y_j\left(t\right) + \varepsilon_i\left(t\right)$, where var$(\varepsilon_i\left(t\right)) = 1/\sigma^{ii}\left(t\right)$. Let $\mathbf{y}_i^k = (y_i^k(t_{k1}), \ldots, y_i^k(t_{km}))$ and $\mathbf{Y}_i = (\mathbf{y}_i^1, \ldots, \mathbf{y}_i^n)'$ be a $nm \times 1$ vector. Let $\mathbf{X}_{-i} =$

$(\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(i-1)}, \mathbf{x}_{(i+1)}, \ldots, \mathbf{x}_{(p)})$ be a $(nm) \times \{(p-1)J_n\}$-dimensional matrix, with $\mathbf{x}_{(j)} = \left(\mathbf{z}_{(j)}^1, \ldots, \mathbf{z}_{(j)}^n\right)'$, and $\mathbf{z}_{(j)}^k = \left(\mathbf{B}(t_{k1})y_j^k(t_{k1}), \ldots, \mathbf{B}(t_{km})y_j^k(t_{km})\right)$. Let $\mathbf{e}_i$ be the residuals from regressing $\mathbf{Y}_i$ to $\mathbf{X}_{-i}$ with

$$\mathbf{e}_i = \mathbf{Y}_i - \mathbf{X}_{-i}\left(\mathbf{X}_{-i}'\mathbf{X}_{-i}\right)^{-1}\mathbf{X}_{-i}'\mathbf{Y}_i.$$

Write $\mathbf{e}_i = (e_{i,ku})_{k=1,u=1}^{n,m}$. Then the initial estimates of $\sigma^{ii}(t)$ at a fixed time point t can be obtained by

$$1/\widehat{\sigma}^{ii}(t) = \sum_{k=1}^{n}\sum_{u=1}^{m} K_h(t_{ku} - t)\, e_{i,ku}^2 / \sum_{k=1}^{n}\sum_{u=1}^{m} K_h(t_{ku} - t),$$

where $K_h(x) = K(x/h)/h$ and $K(x)$ is a kernel function. We assume that the kernel $K(x)$ is a symmetric probability density function on $[-1, 1]$, Lipschitz continuous and has a finite second moment. Kernel smoothing is used here for notation simplicity, but any other smoothing method can also be used here.

Lemma 7. *If the bandwidth h satisfies $nmh^3/\log(nm) \to 0$, $nmh/\log(nm) \to \infty$, and $hN_n \to \infty$, and $\sigma^{ii}(t) \in C_2[0,1]$ for $i = 1, \ldots, p$, then for any $\eta > 0$, there exists a constant c such that $\max_{1 \le i \le p} \sup_{t \in \mathbf{I}} |\widehat{\boldsymbol{\sigma}}^{ii}(t) - \boldsymbol{\sigma}^{ii}(t)| \le c\sqrt{\frac{log(nm)N_n}{nm}}$ holds with probability approaching to 1, as sample size $n \to \infty$.*

Proof: Let $1/\widetilde{\sigma}^{ii}(t) = \sum_{k=1}^{n}\sum_{u=1}^{m} K_h(t_{ku} - t)\,\varepsilon_i^2(t_{ku}) / \sum_{k=1}^{n}\sum_{u=1}^{m} K_h(t_{ku} - t)$ be the Nadaraya-Watson estimator of $1/\sigma^{ii}(t)$ using the true noises $\{\varepsilon_i(t_{ku})\}$ instead. Then by the uniform convergence result on the Nadaraya-Watson estimator (Thm. 6.5. Fan and Yao 2003), one has

$$\sup_{t \in [0,1]} \left|1/\widetilde{\sigma}^{ii}(t) - 1/\sigma^{ii}(t)\right| = O_p\left(\sqrt{\frac{\log(nm)}{nmh}}\right).$$

On the other hand,

$$\sup_{t\in[0,1]} \left|1/\widehat{\sigma}^{ii}(t) - 1/\widetilde{\sigma}^{ii}(t)\right|$$

$$\begin{aligned}
&= \sup_{t\in[0,1]} \left|\sum_{k=1}^{n}\sum_{u=1}^{m} K_h(t_{ku}-t)\left[e_{i,ku}^2 - \varepsilon_i^2(t_{ku})\right] / \sum_{k=1}^{n}\sum_{u=1}^{m} K_h(t_{ku}-t)\right| \\
&= \sup_{t\in[0,1]} \left|\sum_{k=1}^{n}\sum_{u=1}^{m} K_h(t_{ku}-t)\left[e_{i,ku} - \varepsilon_i(t_{ku})\right]\left[e_{i,ku} + \varepsilon_i(t_{ku})\right] / \sum_{k=1}^{n}\sum_{u=1}^{m} K_h(t_{ku}-t)\right| \\
&\le \sup_{t\in[0,1]} \sqrt{\frac{\sum_{k=1}^{n}\sum_{u=1}^{m} K_h(t_{ku}-t)\left[e_{i,ku} - \varepsilon_i(t_{ku})\right]^2}{\sum_{k=1}^{n}\sum_{u=1}^{m} K_h(t_{ku}-t)}}\sqrt{\frac{\sum_{k=1}^{n}\sum_{u=1}^{m} K_h(t_{ku}-t)\left[e_{i,ku} + \varepsilon_i(t_{ku})\right]^2}{\sum_{k=1}^{n}\sum_{u=1}^{m} K_h(t_{ku}-t)}} \\
&= O_p\left(\sqrt{N_n^{-2(q+1)} + \frac{N_n}{nm}}\right),
\end{aligned}$$

where the last step is by the property of the Nadaraya-Watson estimator and the rate of convergence of the polynomial spline estimator involved in the definition of $\mathbf{e}_i$. Therefore together with condition C3 and the conditions on h and N_n, one has,

$$\sup_{t\in[0,1]} \left|1/\widehat{\sigma}^{ii}(t) - 1/\sigma^{ii}(t)\right| = O_p\left(\sqrt{\frac{log(nm)N_n}{nm}}\right).$$

D. Proof of Theorem 1

We first consider $\widetilde{\boldsymbol{\beta}}^*$ that minimizes the objective function which defines the oracle estimator in section 4 of the main text, but with the true $\sigma^{ii}(\cdot)$ instead of $\widehat{\sigma}^{ii}(\cdot)$. Then one can write $\widetilde{\boldsymbol{\beta}}^*$ as

$$\widetilde{\boldsymbol{\beta}}^* = \left(\boldsymbol{\chi}_{n,0}^T\boldsymbol{\chi}_{n,0}\right)^{-1}\boldsymbol{\chi}_{n,0}^T\mathcal{Y}_n.$$

For any $t \in [0,1]$, let

$$\widetilde{\rho}^{*}_{ij}(t) = \mathbf{B}^{T}_{ij}(t)\,\widetilde{\boldsymbol{\beta}}^{*}_{ij}.$$

Furthermore, by Lemma 4, there exist spline functions $g^{(o)}_{ij} \in G^{(o)}_{ij}$, such that $\left\|\rho^{ij} - g^{(o)}_{ij}\right\|_{\infty} \leq c\left(N_n^{-(q+1)}\right)$ for some constant c that does not depend on n. Let

$$\widetilde{m}_i\left(t_{ku}\right) = \sum_{j\neq i}^{p} g^{(o)}_{ij}(t_{ku})\sqrt{\frac{\sigma^{jj}(t_{ku})}{\sigma^{ii}(t_{ku})}}y^{k}_{j}(t_{ku}),$$

and

$$m_i\left(t_{ku}\right) = \sum_{j\neq i}^{p} \rho_{ij}(t_{ku})\sqrt{\frac{\sigma^{jj}(t_{ku})}{\sigma^{ii}(t_{ku})}}y^{k}_{j}(t_{ku}),\ \varepsilon_i\left(t_{ku}\right) = y^{k}_{i}(t_{ku}) - m_i\left(t_{ku}\right).$$

Let $\widetilde{\mathbf{m}}_{iu} = \sqrt{\frac{w_{iu}}{nm}}(\widetilde{m}_i(t_{1u}),\ldots,\widetilde{m}_i(t_{nu}))'$, $\widetilde{\mathbf{m}}_i = (\widetilde{\mathbf{m}}'_{i1},\ldots,\widetilde{\mathbf{m}}'_{im})'$, and $\widetilde{\boldsymbol{M}} = (\widetilde{\mathbf{m}}'_1,\ldots,\widetilde{\mathbf{m}}'_p)'$. One defines $\mathbf{M}$ and $\mathbf{E}$ similarly as $\widetilde{\boldsymbol{M}}$, but using $m_i\left(t_{ku}\right)$ and $\varepsilon_i\left(t_{ku}\right)$, respectively. Then one has,

$$\begin{aligned}
&\widetilde{\rho}^{*}_{ij}(t) - \rho^{ij}(t)\\
=\ &\mathbf{B}^{T}_{ij}(t)\left(\boldsymbol{\chi}^{T}_{n}\boldsymbol{\chi}_{n}\right)^{-1}\boldsymbol{\chi}^{T}_{n}\mathbf{E} + \mathbf{B}^{T}_{ij}(t)\left(\boldsymbol{\chi}^{T}_{n}\boldsymbol{\chi}_{n}\right)^{-1}\boldsymbol{\chi}^{T}_{n}\left(\mathbf{M} - \widetilde{\boldsymbol{M}}\right) + g_{ij}(t) - \rho^{ij}(t)\\
=\ &I(t) + II(t) + III(t).
\end{aligned}$$

For $I(t)$, the Cauchy-Schwartz inequality gives that

$$|I(t)| \leq \sqrt{\mathbf{B}^{T}_{ij}(t)\left(\boldsymbol{\chi}^{T}_{n}\boldsymbol{\chi}_{n}\right)^{-1}\mathbf{B}_{ij}(t)}\sqrt{\mathbf{E}^{T}\boldsymbol{\chi}_{n}\left(\boldsymbol{\chi}^{T}_{n}\boldsymbol{\chi}_{n}\right)^{-1}\boldsymbol{\chi}^{T}_{n}\mathbf{E}}.$$

Then condition (C1) and Lemmas 3 entail that there exists a constant $c > 0$ such that

$$|I(t)| \leq \frac{cN_n}{nm}\sqrt{\mathbf{B}^{T}_{ij}(t)\,\mathbf{B}_{ij}(t)}\sqrt{\mathbf{E}^{T}\boldsymbol{\chi}_{n,0}\boldsymbol{\chi}^{T}_{n,0}\mathbf{E}},$$

in which $\sqrt{\mathbf{E}^T \boldsymbol{\chi}_{n,0} \boldsymbol{\chi}_{n,0}^T \mathbf{E}} = \sqrt{\sum_{i=1}^p \sum_{j\neq i} \sum_{h=1}^{J_n} \left(\sum_{k=1}^n \sum_{u=1}^m B_h(t_{ku}) \sqrt{\frac{\sigma^{jj}(t_{ku})}{\sigma^{ii}(t_{ku})}} y_j^k(t_{ku}) \varepsilon_i\left(t_{ku}\right) \right)^2} = O_p\left(\sqrt{nm}\right)$. On the other hand, $\sqrt{\mathbf{B}_{ij}^T\left(t\right)\mathbf{B}_{ij}\left(t\right)} \leq \sqrt{N_n + q + 1}$, since B-spline bases are upper bounded by 1. Thus one has

$$\sup_{0<t<1} |I(t)| = O_p\left(\frac{N_n^{3/2}}{\sqrt{nm}}\right). \tag{S0.4}$$

Similarly, one can show that

$$\sup_{0<t<1} |II(t)| \leq \frac{cN_n}{nm} \sqrt{\mathbf{B}_{ij}^T\left(t\right)\mathbf{B}_{ij}\left(t\right)} \sqrt{\left(\mathbf{M} - \widetilde{\boldsymbol{M}}\right)^T \boldsymbol{\chi}_n \boldsymbol{\chi}_n^T \left(\mathbf{M} - \widetilde{\boldsymbol{M}}\right)} = O_p\left(\frac{N_n^{-(1+2q)/2}}{\sqrt{nm}}\right). \tag{S0.5}$$

Lastly,

$$\sup_{0<t<1} |III(t)| \leq \sup_{0<t<1} |g_{ij}(t) - \rho^{ij}(t)| \leq c\left(N_n^{-1}\right). \tag{S0.6}$$

Therefore, $\sup_{0<t<1}\left|\widetilde{\rho}_{ij}^*(t) - \rho^{ij}(t)\right| = O_p\left(N_n^{-1} + \frac{N_n^{3/2}}{\sqrt{nm}}\right)$ by combining (S0.4), (S0.5) and (S0.6), and condition (C8). Let $\widetilde{\boldsymbol{\beta}}$ be the spline coefficients for the oracle estimator and $\widetilde{\rho}_{ij}(t) = \mathbf{B}_{ij}^T\left(t\right)\widetilde{\boldsymbol{\beta}}_{ij}$. By Lemma 5 and the convexity of the objective function, the argmax continuous mapping theorem ensures that

$$\sup_{0<t<1}\left|\widetilde{\rho}_{ij}(t) - \rho^{ij}(t)\right| = O_p\left(N_n^{-1} + \frac{N_n^{3/2}}{\sqrt{nm}}\right). \tag{S0.7}$$

E. Proof of Theorem 2

Let $\alpha_n = N_n^{-1} + \sqrt{N_n/nm}$. For any $g(t) \in M_n$ with $\|g - \widetilde{\rho}\|_2 = c\alpha_n$, write $g = \widetilde{\rho} + \delta_n$ with $\|\delta_n\|_2 = c\alpha_n$ for a constant $c > 0$. Denote the corresponding spline coefficients for $g(t), \widetilde{\rho}\left(t\right)$, and $\delta_n\left(t\right)$ as $\left\{\beta_{ij}^h\right\}$, $\left\{\widetilde{\beta}_{ij}^h\right\}$,and $\left\{u_{ij}^h\right\}$ respectively. Then one has

$$PL\left(g\right) = \frac{1}{2}\left\|\mathcal{Y}_n - \mathcal{X}_n\left(\widetilde{\boldsymbol{\beta}} + \mathbf{u}\right)\right\|^2 + \sum_{i<j}^p \sum_{h=1}^{N_n+1} \lambda_n \tau_h^{ij} \|\widetilde{\boldsymbol{\gamma}}_h^{ij} + \widetilde{\mathbf{u}}_h^{ij}\|,$$

and

$$
\begin{aligned}
& PL(g) - PL(\widetilde{\rho}) \\
= \ & \frac{1}{2}\left\|\mathcal{Y}_n - \mathcal{X}_n\left(\widetilde{\boldsymbol{\beta}} + \mathbf{u}\right)\right\|^2 - \frac{1}{2}\left\|\mathcal{Y}_n - \mathcal{X}_n\widetilde{\boldsymbol{\beta}}\right\|^2 + \sum_{i<j}^{p}\sum_{h=1}^{N_n+1}\lambda_n\tau_h^{ij}\left(\|\widetilde{\boldsymbol{\gamma}}_h^{ij} + \overline{\mathbf{u}}_h^{ij}\| - \|\widetilde{\boldsymbol{\gamma}}_h^{ij}\|\right) \\
= \ & I + II,
\end{aligned}
$$

in which by similar arguments as in the proof of Theorem 2 in Xue (2009), there exists a constant $C > 0$, such that $2I = \left\|\mathcal{Y}_n - \mathcal{X}_n\left(\widetilde{\boldsymbol{\beta}} + \mathbf{u}\right)\right\|^2 - \left\|\mathcal{Y}_n - \mathcal{X}_n\widetilde{\boldsymbol{\beta}}\right\|^2 \geq C\alpha_n^2$. For the second term, if $\|\widetilde{\boldsymbol{\gamma}}_h^{ij}\| \neq 0$, then $\tau_h^{ij} \approx 1/\|\widetilde{\boldsymbol{\gamma}}_h^{ij}\|$, and $\|\widetilde{\boldsymbol{\gamma}}_h^{ij} + \overline{\mathbf{u}}_h^{ij}\| - \|\widetilde{\boldsymbol{\gamma}}_h^{ij}\| \approx \left(\overline{\mathbf{u}}_h^{ij}\right)^T\widetilde{\boldsymbol{\gamma}}_h^{ij}/2\|\widetilde{\boldsymbol{\gamma}}_h^{ij}\|$.Therefore, $\lambda_n\tau_h^{ij}\left(\|\widetilde{\boldsymbol{\gamma}}_h^{ij} + \overline{\mathbf{u}}_h^{ij}\| - \|\widetilde{\boldsymbol{\gamma}}_h^{ij}\|\right) \approx \lambda_n\left(\overline{\mathbf{u}}_h^{ij}\right)^T\widetilde{\boldsymbol{\gamma}}_h^{ij}/\left(2\|\widetilde{\boldsymbol{\gamma}}_h^{ij}\|^2\right) \leq \lambda_n\|\overline{\mathbf{u}}_h^{ij}\|/\left(2\|\widetilde{\boldsymbol{\gamma}}_h^{ij}\|\right) = o_p\left(\alpha_n^2/N_n\right)$, by Condition C8. If $\|\widetilde{\boldsymbol{\gamma}}_h^{ij}\| = 0$, then $\lambda_n\tau_h^{ij}\left(\|\widetilde{\boldsymbol{\gamma}}_h^{ij} + \overline{\mathbf{u}}_h^{ij}\| - \|\widetilde{\boldsymbol{\gamma}}_h^{ij}\|\right) = \lambda_n\tau_h^{ij}\|\overline{\mathbf{u}}_h^{ij}\| \geq 0$. Thus $II = o_p\left(\alpha_n^2\right)$. Therefore, when n is sufficiently large, for any $\varepsilon > 0$, there exists a sufficiently large $c > 0$ such that

$$
P\left(\inf_{g\in M_n, \|g-\widetilde{\rho}\|=c\alpha_n} PL(g) \geq PL(\widetilde{\rho})\right) \geq 1 - \varepsilon.
$$

Hence there exists a minimizer $\widehat{\rho} \in M_n$ in a neighborhood of $\widetilde{\rho}$ with $\|\widehat{\rho} - \widetilde{\rho}\| = O_p(\alpha_n)$. Together with Theorem 1, one has $\|\widehat{\rho} - \rho\| = O_p(\alpha_n)$.

F. Proof of Theorem 3

Let $c(\beta) = \frac{1}{2nm}\sum_{i=1}^{p}\sum_{k=1}^{n}\sum_{u=1}^{m}\left(y_i^k(t_u) - \sum_{j\neq i}^{p}\sum_{h=1}^{J_n}\beta_h^{ij}B_h^{ij}(t_u)\sqrt{\frac{\widehat{\sigma}^{jj}(t_u)}{\widehat{\sigma}^{ii}(t_u)}}y_j^k(t_u)\right)^2$, $c_h^{ij}(\beta) = \partial c(\beta)/\partial\beta_h^{ij}$, and $\overline{c}_h^{ij}(\beta) = \left(c_h^{ij}(\beta), \ldots, c_{h+p}^{ij}(\beta)\right)$. By the KKT condition, $\widehat{\beta}$ is the solution of the penalized minimization problem if and only if

$$c_k^{ij}\left(\widehat{\beta}\right)+\sum_{s=\max(k-p,1)}^{\min(k,N_n+1)}\frac{\lambda_n\tau_s^{ij}}{\left\|\widehat{\gamma}_s^{ij}\right\|}\widehat{\beta}_k^{ij} = 0, \quad if \quad \left\|\widehat{\gamma}_s^{ij}\right\|\neq 0 \text{ for } \max(k-p,1)\leq s\leq \min(k,N_n+1),$$
$$\left\|\overline{\mathbf{c}}_k^{ij}\left(\widehat{\beta}\right)\right\|_2 \leq \lambda_n\tau_k^{ij}, \quad if \quad \left\|\widehat{\gamma}_k^{ij}\right\|=0. \tag{S0.8}$$

Let $\beta^{ij}=\left(\beta_1^{ij},\ldots,\beta_{J_n}^{ij}\right)^T$, and $\beta_{J_{ij}}^{ij}=\left(\beta_k^{ij},k\in J_{ij}\right)^T$. Define $\beta_{J_{ij}^C}^{ij}$ similarly. Let $\widehat{\beta}=\left(\widehat{\beta}^{ij},1\leq i<j\leq p\right)$ such that for each $\widehat{\beta}^{ij}$ with $\widehat{\beta}_{J_{ij}^C}^{ij}=0$ and $\widehat{\beta}_{J_{ij}}^{ij}$ solving

$$c_k^{ij}\left(\beta\right)+\sum_{s=\max(k-p,1)}^{\min(k,N_n+1)}\frac{\lambda_n\tau_s^{ij}}{\left\|\widehat{\gamma}_s^{ij}\right\|}\beta_k^{ij}=0, \tag{S0.9}$$

for $k\in J_{ij}$. Write $\widehat{\beta}^{(A)}=\left\{\widehat{\beta}_{J_{ij}}^{ij},1\leq i<j\leq p\right\}$. Then

$$\widehat{\beta}^{(A)}=\left(\boldsymbol{\chi}_{n,0}^T\boldsymbol{\chi}_{n,0}\right)^{-1}\left(\boldsymbol{\chi}_{n,0}^T\mathcal{Y}_n+\mathbf{W}_n\right),$$

where $\mathbf{W}_n=\left\{\sum_{s=\max(k-p,1)}^{\min(k,N_n+1)}\lambda_n\tau_s^{ij}\beta_k^{ij}/\left\|\widehat{\gamma}_s^{ij}\right\|,k\in J_{ij}\right\}$. Theorem 2 entails that elements in $\mathbf{W}_n$: $w_k^{ij}=\sum_{s=\max(k-p,1)}^{\min(k,N_n+1)}\lambda_n\tau_s^{ij}/\left\|\widehat{\gamma}_s^{ij}\right\|=O_p\left(\lambda_n\right)$. Therefore, together with Lemma 3, one has, $\left\|\widehat{\beta}^{(A)}-\widetilde{\beta}^{(A)}\right\|=\sqrt{\mathbf{W}_n^T\left(\boldsymbol{\chi}_{n,0}^T\boldsymbol{\chi}_{n,0}\right)^{-2}\mathbf{W}_n}=O_p\left(\lambda_nN_n/nm\right)$, and $\left\|\mathbf{c}_h^{ij}\left(\widehat{\beta}\right)-\mathbf{c}_h^{ij}\left(\widetilde{\beta}\right)\right\|=O_p\left(\lambda_n\right)$. Then Lemma 6 and condition (C9) entail that

$$P\left(\max_{1\leq i<j\leq p,k\in J_{ij}^c}\left\|\overline{\mathbf{c}}_k^{ij}\left(\widehat{\beta}\right)\right\|_2\geq\lambda_n\tau_k^{ij}\right)\to 0.$$

Therefore, $\widehat{\beta}$ satisfies the KKT condition, and is the solution of the adaptive Lasso objective function. It is clear from the definition of $\widehat{\beta}^{ij}$ that $\widehat{\rho}^{ij}\left(t\right)=0$ for $t\in E^{ij}=[e_1^{ij},e_2^{ij}]$. Now let $\widehat{\rho}^{ij}=\left(\widehat{\beta}^{ij}\right)^TB$ be the corresponding estimator of the partial correlation functions. We now show that $\sup_{0\le t\le 1}\left|\widehat{\rho}^{ij}(t)-\rho^{ij}(t)\right| = O_p\left(N_n^{-1}\right)$. One notices that

$$\widehat{\rho}^{ij}(t)-\widetilde{\rho}^{ij}(t)=\left(\mathbf{B}^{(A)}(t)\right)^T\left(\left(\mathcal{X}_n^{(A)}\right)^T\mathcal{X}_n^{(A)}+W_n^{(A)}\right)^{-1}W_n^{(A)}\widetilde{\beta}^{(A)},$$

and by Lemma 3

$$\begin{aligned}\sup_{0<t<1}\left|\widehat{\rho}^{ij}(t)-\widetilde{\rho}^{ij}(t)\right| &\le \frac{N_n}{nm}\lambda_n\sup_{0<t<1}\left|\mathbf{B}_{J_{ij}}^T(t)\,\widetilde{\beta}_{J_{ij}}^{ij}\right|=\frac{N_n}{nm}\lambda_n\sup_{0<t<1}\left|\widetilde{\rho}^{ij}(t)\right| \\ &= O_p\left(\frac{N_n}{nm}\lambda_n\right)=o_p\left(\frac{N_n^{3/2}}{\sqrt{nm}}\right).\end{aligned}$$

Then Theorem 3 follows from the triangular inequality and condition (C8). □

Bibliography

[1] de Boor, C. (2001). *A Practical Guide to Splines*, New York: Springer.

[2] Devore, R. A., & Lorentz, G. G. (1993). *Constructive Approximation*, Berlin Heidelberg: Springer-Varlag.

[3] Fan & Yao (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer.

[4] Xue, L. (2009). Variable selection in additive models. *Statistica Sinica* **19**, 1281-1296.

[5] Xue, L. & Qu, A. (2012). Variable selection in high-dimensional varying-coefficient models with global optimality. *J. Mach. Learn. Res.* **13**, 1973-1998.

[6] Xue, L. & Yang, L. (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica* **16**, 1423-1446.