Web-based Supplementary Material for "Effective Inference Procedures for Partially Nonlinear Models" by Runze Li and Lei Nie

We first present the regularity conditions for Theorems 1 to 4 in this Web Appendix.

Regularity Conditions

(A) The random variable U has a bounded support $\mathcal{U}$. Its density function $f(u)$ is Lipschitz continuous and bounded from 0 on its support.

(B) The true unknown smoothing function $\alpha_0(u)$ has a continuous second derivative.

(C) $K(u)$ is a positive, bounded, and symmetric function with compact support. Furthermore, $K(u)$ satisfies the Lipschitz condition. The functions $u^3K(u)$ and $u^3K'(u)$ are bounded and $\int u^4K(u)\,du < \infty$.

(D) $nh^8 \to 0$ and $nh^2/\{\ln(h)\}^2 \to \infty$.

(E) For any $\mathbf{x}$, $g(\mathbf{x}, \boldsymbol{\beta})$ is a continuous function of $\boldsymbol{\beta}$ and the second derivatives of $g(\mathbf{x}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ are continuous, $\boldsymbol{\beta} \in \mathcal{B}$, where $\mathcal{B}$ is a compact set.

(F) Let d be the dimension of $\boldsymbol{\beta}$, and

$$g'(\mathbf{x}_i, \boldsymbol{\beta}) = [\partial g(\mathbf{x}_i, \boldsymbol{\beta})/\partial \boldsymbol{\beta}]_{d\times 1}, \text{ and } g''(\mathbf{x}_i, \boldsymbol{\beta}) = \left[\partial^2 g(\mathbf{x}_i, \boldsymbol{\beta})/\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T\right]_{d\times d},$$

and $\text{Vech}\{\text{g}''(\mathbf{x}, \boldsymbol{\beta})\}$ is the $d \times (d+1)/2$-vector of all second derivatives of $g(\mathbf{x}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. $\text{E}\{g'(\mathbf{x}, \boldsymbol{\beta})\}^{\otimes 2}$, $\text{E}[\text{E}\{g'(\mathbf{x}, \boldsymbol{\beta})|U\}^{\otimes 2}]$, and $\text{E}(\text{E}[\{\text{Vech}\{\text{g}''(\mathbf{x}, \boldsymbol{\beta})\}|\text{U}]^{\otimes 2})$ are bounded in a neighborhood of $\boldsymbol{\beta}_0$.

(G) $\text{E}\{\|g'(\mathbf{x}, \boldsymbol{\beta})\|^4\} < \infty$, $\text{E}[\|\text{Vech}\{\text{g}''(\mathbf{x}, \boldsymbol{\beta})\}\|^4] < \infty$.

(H) $\|\text{Vech}\{g''(\mathbf{x}, \boldsymbol{\beta})\}\| \le B(\mathbf{x})$ for all $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_0$ and $\text{E}\{\|B(\mathbf{x})\|^4\} < \infty$.

Let

$$D_u = \begin{pmatrix} 1 & \frac{u_1-u}{n} \\ . & . \\ . & . \\ . & . \\ 1 & \frac{u_n-u}{n} \end{pmatrix}_{n\times 2},$$

and $W_u = \text{diag}\{K_h(u_1-u), \cdots, K_h(u_n-u)\}$. By definition of S_h,

$$S_h = \begin{pmatrix} (1,0)(D_{u_1}^T W_{u_1} D_{u_1})^{-1} D_{u_1}^T W_{u_1} \\ . \\ . \\ . \\ (1,0)(D_{u_n}^T W_{u_n} D_{u_n})^{-1} D_{u_n}^T W_{u_n} \end{pmatrix}_{n\times n}.$$

Let $c_n = \{\frac{-\ln(h)}{nh}\}^{1/2} + h^2$,

$$\mathbf{g}'(\boldsymbol{\beta}) = (g'(\mathbf{x}_1, \boldsymbol{\beta}), \cdots, g'(\mathbf{x}_n, \boldsymbol{\beta}))^T,$$

and

$$\xi_n = n^{-1} \sum_{i=1}^{n} [g'(\mathbf{x}_i; \boldsymbol{\beta}_0) - E\{g'(\mathbf{x}; \boldsymbol{\beta}_0)|U = u_i\}]\varepsilon_i.$$

The following lemma is used in the proof of Theorems 1 and 2 repeatedly.

Lemma 1. Under Conditions (A) — (H), it follows that

$$\frac{1}{n}\mathbf{g}'(\boldsymbol{\beta}_0)^T (I_n - S_h)^T (I_n - S_h)\mathbf{g}'(\boldsymbol{\beta}_0) = \mathbf{A}\{1 + o_p(1)\}, \quad \text{(A.1)}$$

$$\frac{1}{n}\mathbf{g}'(\boldsymbol{\beta}_0)^T (I_n - S_h)^T (I_n - S_h)\{\mathbf{y} - \mathbf{g}(\boldsymbol{\beta}_0)\} = \xi_n + O_p(c_n^2). \quad \text{(A.2)}$$

Lemma 1 can be proved by using Proposition 4 in Marc and Silverman (1982) and related techniques in the proofs of Lemmas 7.1 to 7.4 in Fan and Huang (2005).

Proof of Theorem 1. Let $Q'_j(\boldsymbol{\beta})$ denote the j-th component of $Q'(\boldsymbol{\beta})$, and $Q''_j(\boldsymbol{\beta})$ be the j-row of $Q''(\boldsymbol{\beta})$. Using Taylor's expansion, for $j = 1, \cdots, d$,

$$0 = Q'_j(\hat{\boldsymbol{\beta}}) = Q'_j(\boldsymbol{\beta}_0) + Q''_j(\boldsymbol{\beta}^*_j)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \tag{A.3}$$

where $\boldsymbol{\beta}^*_j$ lies between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. Under conditions (A)-(H), it can be shown that

$$\frac{1}{2n} Q''_j(\boldsymbol{\beta}^*_j) = A_j\{1 + o_p(1)\},$$

in probability, where A_j is the j-row of $\mathbf{A}$. Using (A.2), it follows that

$$n^{-1}Q'(\boldsymbol{\beta}_0) = -2n^{-1}\mathbf{g}'(\boldsymbol{\beta}_0)^T(I_n - S_h)^T(I_n - S_h)\{\mathbf{y} - \mathbf{g}(\boldsymbol{\beta}_0)\} = -2\xi_n + O_p(c_n^2). \tag{A.4}$$

Thus,

$$\sqrt{n}\mathbf{A}\{1 + o_p(1)\}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \sqrt{n}\{\xi_n + O_p(c_n^2)\} = \sqrt{n}\xi_n + o_P(1),$$

as $\sqrt{n}c_n^2 \to 0$ by Condition (D). Note that,

$$\xi_n = n^{-1}\sum_{i=1}^{n}[g'(\mathbf{x}_i; \boldsymbol{\beta}_0) - E\{g'(\mathbf{x}; \boldsymbol{\beta}_0)|U = u_i\}]\varepsilon_i.$$

Using the Slutsky theorem and the central limit theorem, it follows that

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2\mathbf{A}^{-1}).$$

Proof of Theorem 2. Let $\mathbf{g}(\boldsymbol{\beta}) = (g(\mathbf{x}_1, \boldsymbol{\beta}), \cdots, g(\mathbf{x}_n, \boldsymbol{\beta}))^T$. Note that

$$\hat{\boldsymbol{\beta}}_L - \boldsymbol{\beta}_0 = \{\mathbf{g}'(\hat{\boldsymbol{\beta}}_I)^T(I_n - S_h)^T(I_n - S_h)\mathbf{g}'(\hat{\boldsymbol{\beta}}_I)\}^{-1}\mathbf{g}'(\hat{\boldsymbol{\beta}}_I)^T(I_n - S_h)^T(I_n - S_h)\{\mathbf{z} - \mathbf{g}'(\hat{\boldsymbol{\beta}}_I)\boldsymbol{\beta}_0\},$$

where $\mathbf{z} = (z_1, \cdots, z_n)^T$ (defined in Section 2.2) and $\mathbf{g}'(\hat{\boldsymbol{\beta}}_I) = (g'(\mathbf{x}_1; \hat{\boldsymbol{\beta}}_I), \cdots, g'(\mathbf{x}_n; \hat{\boldsymbol{\beta}}_I))^T$. It has been shown in the earlier version of this paper (Li and Nie, 2006) that

$$\frac{1}{n}\mathbf{g}'(\hat{\boldsymbol{\beta}}_I)^T(I_n - S_h)^T(I_n - S_h)\mathbf{g}'(\hat{\boldsymbol{\beta}}_I) = \mathbf{A}\{1 + o_p(1)\}. \tag{A.5}$$

We next show that

$$\frac{1}{\sqrt{n}}\mathbf{g}'(\hat{\boldsymbol{\beta}}_I)^T(I_n - S_h)^T(I_n - S_h)(\mathbf{z} - \mathbf{g}'(\hat{\boldsymbol{\beta}}_I)\boldsymbol{\beta}_0) = \sqrt{n}\xi_n + o_P(1). \tag{A.6}$$

Using the definition of $\mathbf{z}$, we have

$$\mathbf{z} - \mathbf{g}'(\hat{\boldsymbol{\beta}}_I)\boldsymbol{\beta}_0 = \mathbf{y} - \mathbf{g}(\hat{\boldsymbol{\beta}}_I) + \mathbf{g}'(\hat{\boldsymbol{\beta}}_I)(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0)$$

and it follows by (A.2) that

$$\frac{1}{\sqrt{n}}\mathbf{g}'(\boldsymbol{\beta}_0)^T(I_n - S_h)^T(I_n - S_h)\{\mathbf{y} - \mathbf{g}(\boldsymbol{\beta}_0)\} = \sqrt{n}\xi_n + O_P(\sqrt{n}c_n^2) = \sqrt{n}\xi_n + o_P(1).$$

Thus, to establish (A.6), it is enough to show that

$$\frac{1}{\sqrt{n}}\mathbf{g}'(\boldsymbol{\beta}_0)^T(I_n - S_h)^T(I_n - S_h)\{\mathbf{g}(\boldsymbol{\beta}_0) - \mathbf{g}(\hat{\boldsymbol{\beta}}_I) + \mathbf{g}'(\hat{\boldsymbol{\beta}}_I)(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0)\} = o_p(1), \tag{A.7}$$

and

$$\frac{1}{\sqrt{n}}\{\mathbf{g}'(\hat{\boldsymbol{\beta}}_I) - \mathbf{g}'(\boldsymbol{\beta}_0)\}^T(I_n - S_h)^T(I_n - S_h)\{\mathbf{z} - \mathbf{g}'(\hat{\boldsymbol{\beta}}_I)\boldsymbol{\beta}_0\} = o_p(1). \tag{A.8}$$

By straightforward calculation, the left-hand side of (A.7) is of the order

$$O_P(\sqrt{n}\|\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0\|^2) = O_P(1/\sqrt{n})$$

as $\|\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0\| = O_P(n^{-1/2})$. Furthermore, the left-hand side of (A.8) is of the order $O_P(c_n\|\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_0\|) = O_P(c_n/\sqrt{n})$. Thus, (A.6) holds.

Using (A.5), (A.6), it follows

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_L - \boldsymbol{\beta}_0) = \mathbf{A}\{1 + o_P(1)\}^{-1}\{\sqrt{n}\xi_n + o_P(1)\}.$$

By the Slutsky Theorem and the central limit theorem, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_L - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2\mathbf{A}^{-1}).$$

This completes the proof of Theorem 2.

Equivalence between algorithm (2.12) and Fisher scoring algorithm

We here demonstrate algorithm (2.12) is equivalent to using the Fisher scoring algorithm to minimize $Q(\boldsymbol{\beta})$ in (2.5). The Newton-Raphson algorithm to minimize $Q(\boldsymbol{\beta})$ in (2.5) is to iteratively compute

$$\hat{\boldsymbol{\beta}}^{(m+1)} = \hat{\boldsymbol{\beta}}^{(m)} - Q''(\hat{\boldsymbol{\beta}}^{(m)})^{-1} Q'(\hat{\boldsymbol{\beta}}^{(m)}).$$

Note that

$$E\{Q''(\boldsymbol{\beta}_0)\} = 2E\{\mathbf{g}'(\boldsymbol{\beta}_0)^T (I - S_h)^T (I - S_h) \mathbf{g}'(\boldsymbol{\beta}_0)\} \hat{=} I(\boldsymbol{\beta}_0)$$

which corresponds to the Fisher information matrix. Thus, the Fisher scoring algorithm is to iteratively compute

$$\hat{\boldsymbol{\beta}}^{(m+1)} = \hat{\boldsymbol{\beta}}^{(m)} - \hat{I}^{-1}(\hat{\boldsymbol{\beta}}^{(m)}) Q'(\hat{\boldsymbol{\beta}}^{(m)}),$$

where $\hat{I}(\hat{\boldsymbol{\beta}}^{(m)}) = 2\{\mathbf{g}'(\hat{\boldsymbol{\beta}}^{(m)})^T (I - S_h)^T (I - S_h) \mathbf{g}'(\hat{\boldsymbol{\beta}}^{(m)})\}$. Thus the corresponding Fisher score algorithm is to iteratively calculate

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{(m+1)} &= \hat{\boldsymbol{\beta}}^{(m)} + \left\{\mathbf{g}'(\hat{\boldsymbol{\beta}}^{(m)})^T (I - S_h)^T (I - S_h) \mathbf{g}'(\hat{\boldsymbol{\beta}}^{(m)})\right\}^{-1} \\ &\quad \times \mathbf{g}'(\hat{\boldsymbol{\beta}}^{(m)})^T (I - S_h)^T (I - S_h) \left\{y - \mathbf{g}(\hat{\boldsymbol{\beta}}^{(m)})\right\}, \end{aligned}$$

where $\mathbf{g}(\boldsymbol{\beta}^{(m)}) = (g(\mathbf{x}_1, \boldsymbol{\beta}^{(m)}), \cdots, g(\mathbf{x}_n, \boldsymbol{\beta}^{(m)}))^T$. Since $\mathbf{z}^{(m)} = \mathbf{y} - \mathbf{g}(\hat{\boldsymbol{\beta}}^{(m)}) + \mathbf{g}'(\hat{\boldsymbol{\beta}}^{(m)})^T \hat{\boldsymbol{\beta}}^{(m)}$, it follows that

$$\hat{\boldsymbol{\beta}}^{(m+1)} = \left\{\mathbf{g}'(\hat{\boldsymbol{\beta}}^{(m)})^T (I - S_h)^T (I - S_h) \mathbf{g}'(\hat{\boldsymbol{\beta}}^{(m)})\right\}^{-1} \mathbf{g}'(\hat{\boldsymbol{\beta}}^{(m)})^T (I - S_h)^T (I - S_h) \mathbf{z}^{(m)}.$$

which is (2.12).

Proof of Theorem 4. Let $y_i^* = y_i - g(\mathbf{x}_i; \boldsymbol{\beta}_0)$. Thus,

$$y_i^* = \alpha(u_i) + \varepsilon_i.$$

Let $\tilde{\alpha}^*$ and $\hat{\alpha}^*(\cdot)$ be the estimate of α under H_0 and H_1, respectively. Denote $\mathrm{RSS}^*(H_0) = \sum_{i=1}^n (y_i^* - \tilde{\alpha}^*)^2$ and $\mathrm{RSS}^*(H_1) = \sum_{i=1}^n \{y_i^* - \hat{\alpha}^*(u_i)\}^2$. Define

$$\mathrm{GLRT}_0^* = (n/2)(\mathrm{RSS}^*(H_0) - \mathrm{RSS}^*(H_1))/\mathrm{RSS}^*(H_1)$$

By Theorem 5 of Fan, Zhang and Zhang (2001), it follows that

$$r_K \mathrm{GLRT}_0^* \overset{a}{\sim} \chi^2_{\delta_n}.$$

Note that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_P(n^{-1/2})$. By Lemma 1 and the Taylor expansion, we have

$$\{\mathrm{RSS}(H_1) - \mathrm{RSS}^*(H_1)\} = -n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \mathbf{A} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_P(1),$$

and under H_0, it can be shown by using theory of linear regression that

$$\{\mathrm{RSS}(H_0) - \mathrm{RSS}^*(H_0)\} = -n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \mathbf{A} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_P(1).$$

The proof is completed by noticing that $\mathrm{RSS}^*(H_1)/\mathrm{RSS}(H_1) \to 1$ in probability.

Additional References

Fan, J. and Huang, T. (2005). Profile Likelihood Inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, **11**, 1031-1057.

Li, R. and Nie, L. (2006). Efficient statistical inference procedures for partially nonlinear models and their applications. Available at

http://www.stat.psu.edu/˜rli/research/nlnannTR.pdf

Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **61**, 405–415.