Web-Based Supplementary Materials for

“A Partial Linear Model in the Outcome Dependent Sampling Setting to Evaluate the Effect of Prenatal PCB Exposure on Cognitive Function in Children”

by Haibo Zhou, Guoyou Qin and Matthew P. Longnecker

Web Appendix A: Selection of Knots, Penalty and Smoothing Parameters

In practice, we need to choose the appropriate knots and determine the magnitude of the smoothing parameter q_s. For the selection of knots, we recommend that the knots be placed at equally spaced sample quantiles of the predictor variable. As suggested by Ruppert (2002) and confirmed by our simulation study, 5-10 knots seem quite adequate for most smoothing applications, especially monotonic or unimodal functions. If the nonparametric function has a discontinuity, then it is important to have a knot near it. A detailed discussion of the choice of knots has been given in Ruppert (2002).

The quadratic penalty function is commonly used. However, in some cases, e.g. the nonparametric function has discontinuity, the non-quadratic penalty may be a better choice for P-spline smoothing. Ruppert and Carroll (1997) gave a general δ penalty form $\sum_{t=1}^{T} |\alpha_{r+t}|^{\delta}, \delta > 0$, and pointed out that penalties with δ less than or equal to 1 can outperform a quadratic penalty for the discontinuity function. Otherwise, the quadratic penalty is preferred.

Selecting a suitable value of smoothing parameter q_s is crucial to good curve fitting. In this article, we borrow the idea of Qu and Li (2006) and define the generalized cross-validation (GCV) score as,

$$\text{GCV}(q_s) = -\frac{\frac{1}{n} l_n^*}{(1 - \frac{1}{n}\text{df})^2},$$

where $l_n^* = ppl_n(\eta; q_s) + \frac{1}{2} n q_s \theta^T \Psi \theta$, $\text{df} = \text{trace}\{G(q_s)\}$ is the effective degree of freedom, $G(q_s) = \left(\frac{\partial^2 l_n^*}{\partial\theta\partial\theta^T} - n q_s \Psi \right)^{-1} \frac{\partial^2 l_n^*}{\partial\theta\partial\theta^T}$. Then $\hat{q}_s = \text{argmin}_{q_s} \text{GCV}(q_s)$. In practice, this minimization can be carried out by searching over a grid of q_s values. Similar to Yu and Ruppert (2002) and according to our simulation experience, in this article we select q_s over 30 grid points where the values of $log_{10}(q_s)$ are equally spaced between -6 and 7.

Web Appendix B: Assumptions and Proof of Theorem 1

Some notations are introduced as follows:

The first derivative of $ppl_n(\eta; q_s)$ is denoted as

$$Q_n(\eta; q_s) = \frac{\partial ppl_n(\eta)}{\partial \eta} = \sum_{k=0}^{K} \left\{ \sum_{j=1}^{n_k} u^*(y_{kj}, x_{kj}, z_{kj}; \eta) \right\},$$

where $u^*(y_{kj}, x_{kj}, z_{kj}; \eta) =$

$$\begin{pmatrix} \frac{\partial log f(y_{kj}|x_{kj},z_{kj};\theta)}{\partial\theta} - \frac{\partial Q(x_{kj},z_{kj};\eta)/\partial\theta}{Q(x_{kj},z_{kj};\eta)} - \frac{\partial(v^T h(x_{kj},z_{kj};\eta))/\partial\theta}{1+v^T h(x_{kj},z_{kj};\eta)} - q_s\Psi\theta \\ -I_{\{0<k<K\}} 1_k (\frac{1}{\pi_k} - \frac{n_K}{n_k}\frac{1}{\pi_K}) - \frac{\partial Q(x_{kj},z_{kj};\eta)/\partial\pi}{Q(x_{kj},z_{kj};\eta)} - \frac{\partial v^T h(x_{kj},z_{kj};\eta)/\partial\pi}{1+v^T h(x_{kj},z_{kj};\eta)} \\ -\frac{h(x_{kj},z_{kj};\eta)}{1+v^T h(x_{kj},z_{kj};\eta)} \end{pmatrix},$$

where $I_{\{0<k<K\}}$ is a indicator function, 1_k denotes a $(K-1)$-dimensional vector with the kth component is 1 and the rest are zero.

We further denote $u(y_{kj}, x_{kj}, z_{kj}; \eta) =$

$$\begin{pmatrix} \frac{\partial log f(y_{kj}|x_{kj},z_{kj};\theta)}{\partial\theta} - \frac{\partial Q(x_{kj},z_{kj};\eta)/\partial\theta}{Q(x_{kj},z_{kj};\eta)} - \frac{\partial(v^T h(x_{kj},z_{kj};\eta))/\partial\theta}{1+v^T h(x_{kj},z_{kj};\eta)} \\ -I_{\{0<k<K\}} 1_k (\frac{1}{\pi_k} - \frac{n_K}{n_k}\frac{1}{\pi_K}) - \frac{\partial Q(x_{kj},z_{kj};\eta)/\partial\pi}{Q(x_{kj},z_{kj};\eta)} - \frac{\partial v^T h(x_{kj},z_{kj};\eta)/\partial\pi}{1+v^T h(x_{kj},z_{kj};\eta)} \\ -\frac{h(x_{kj},z_{kj};\eta)}{1+v^T h(x_{kj},z_{kj};\eta)} \end{pmatrix},$$

which do not involve the penalty term. In the following, for brevity, we sometimes use $u_{kj}(\eta)$ and $u^*_{kj}(\eta)$ to denote $u(y_{kj}, x_{kj}, z_{kj}; \eta)$ and $u^*(y_{kj}, x_{kj}, z_{kj}; \eta)$.

To establish the asymptotic properties, we make the following assumptions:

A1) The parameter space Θ is a compact subset of R^d, the true value $\eta_0 = \{\theta_0^T, \pi_0^T, \mathbf{0}^T_{(K-1)\times 1}\}^T$ lies in the interior of Θ.

A2) Assume $u(Y, X, Z; \eta)$ is continuously differentiable for all $\eta \in \Theta$ and $E||u(Y, X, Z; \eta)||^{2+\delta} < \infty$ for some positive constant δ.

A3) Assume $E\{u(Y, X, Z; \eta)u(Y, X, Z; \eta)^T\}$ and $-E\{\frac{\partial}{\partial\eta}u(\eta)\}$ are positive definite at η_0, and $E||\frac{\partial}{\partial\eta}u(Y, X, Z; \eta)||$ is bounded for all $\eta \in \Theta$ where $||.||$ denotes the Euclidean norm.

A4) Assume that $n_k/n \to \rho_k \geq 0$ for $k = 0, \ldots, K$.

A5) The smoothing parameter q_s is assumed to satisfy $q_s = o(1)$ for the consistency of the proposed estimator and satisfy $q_s = o(1/\sqrt{n})$ for the asymptotic normality.

Following is the proof of Theorem 1.

(i) Proof of Consistency

The consistency proof is similar to the proof of Theorem 4.1 in Lehmann (1983). We sketch the outline of the proof in the following. By the law of large numbers and notice of that the $q_s = o(1)$, we have $\frac{1}{n}V_n(\eta; q_s) = \frac{1}{n}\frac{\partial^2 ppl_n(\eta;q_s)}{\partial\eta\partial\eta^T}$ converges to a negative definite $V(\eta)$ in probability, where $V(\eta) = \sum_{k=0}^{K}\rho_k E\left\{\frac{\partial}{\partial\eta}u_{k1}(\eta)\right\}$. By a Taylor expansion around η_0 in a neighborhood of η_0 with the notice of that $V(\eta)$ is a negative definite matrix, we have $\frac{1}{n}ppl_n(\eta_0) > \frac{1}{n}ppl_n(\eta)$ a.s. in a neighborhood of η_0. Thus, $\frac{1}{n}ppl_n(\eta)$ has a local maximum in a small neighborhood of η_0. The consistency is achieved by the smoothness of the likelihood function.

(ii)Proof of Asymptotic Normality

By the law of large numbers and notice of that $q_s = o(1/\sqrt{n})$, we have $\frac{Q_n(\eta;q_s)}{n} \to Q(\eta)$ in probability, where $Q(\eta) = \sum_{k=0}^{K}\rho_k[E\{u_{k1}(\eta)\}]$. When evaluated at η_0, using the similar arguments to Weaver (2001), we can show $Q(\eta_0) = 0$.

To show the asymptotic normality of $(\hat{\theta}^T, \hat{\pi}^T, \hat{v})^T$, we expand $\frac{1}{n}Q_n(\eta; q_s)$ at η_0 such that

$$\frac{1}{n}Q_n(\eta; q_s) = \frac{1}{n}Q_n(\eta_0; q_s) + \frac{1}{n}V_n(\eta_0; q_s)(\eta - \eta_0) + o_p(n^{-1/2}),$$

then we have

$$\sqrt{n}(\hat{\eta} - \eta_0) = -(\frac{1}{n}V_n(\eta_0; q_s))^{-1}\frac{1}{\sqrt{n}}Q_n(\eta_0; q_s) + o_p(1).$$

If the smoothing parameter $q_s = o(1/\sqrt{n})$, then $\frac{1}{\sqrt{n}}nq_s\Psi\theta = \sqrt{n}q_s\Psi\theta = o(1)$. By the central limit theorem, we have $\frac{1}{\sqrt{n}}Q_n(\eta_0; q_s) \to N(\mathbf{0}_{d\times 1}, U(\eta_0))$ in distribution, where

$$U(\eta_0) = \sum_{k=0}^{K}\rho_k[Cov\{u_{k1}(\eta_0)\}],$$

According to that $\frac{1}{n}V_n(\eta; q_s)$ converges to $V(\eta)$ in probability and continuous mapping theorem, we have $\sqrt{n}(\hat{\eta} - \eta_0) \to N(\mathbf{0}_{d\times 1}, \Sigma)$ in distribution. The covariance matrix Σ can be estimated consistently by $\hat{\Sigma} = \hat{V}^{-1}\hat{U}\hat{V}^{-1}$, where $\hat{V} = \frac{1}{n}V_n(\hat{\eta}; q_s)$, and $\hat{U} = \frac{1}{n}\sum_{k=0}^{K}\sum_{j=1}^{n_k}\hat{u}_{kj}^*(\hat{\eta})\hat{u}_{kj}^{*T}(\hat{\eta})$ by the consistent results and the continuous mapping theorem.