BIOMETRICS XXX, 1–??
XXX 2015

DOI: 10.1111/j.1541-0420.2005.00454.x

Web-based Supplementary Materials for "Doubly Robust Estimation of Generalized Partial Linear Models for Longitudinal Data with Dropouts" by Lin, Fu, Qin and Zhu

Huiming Lin[1,4], Bo Fu[2,5], Guoyou Qin[1,4,*], and Zhongyi Zhu[3]

[1]Department of Biostatistics, School of Public Health and Key Laboratory of Public Health Safety, Fudan University, Shanghai, 200032, China

[2] Administrative Data Research Centre for England & Institute of Child Health, University College London, London, UK

[3] Department of Statistics, Fudan University, Shanghai, 200433, China

[4] Collaborative Innovation Center of Social Risks Governance in Health

[5] Centre for Biostatistics, Institute of Population Health, The University of Manchester, Manchester, UK

*email: gyqin@fudan.edu.cn

This paper has been submitted for consideration for publication in *Biometrics*

1. Web Appendix A: Proof of the double robustness of the proposed estimator

Recall the proposed estimating equation

$$\sum_{i=1}^{n} U_i = \sum_{i=1}^{n} \{\dot{\mu}_i \Sigma_i^{-1} W_i (Y_i - \mu_i) + \dot{\mu}_i \Sigma_i^{-1} (I_i - W_i) \tilde{E}(Y_i - \mu_i)\} = 0$$

When the LCM condition holds but the missing probability model is wrong,

$$\begin{aligned}
E(U_i) =& E\left(\dot{\mu}_i \Sigma_i^{-1} \begin{bmatrix} E(Y_{i1} - \mu_{i1}|X_i) \\ E(\frac{R_{i2}}{\pi_{i2}}[Y_{i2} - \mu_{i2}]|\tilde{Y}_i^2, X_i) + E([1 - \frac{R_{i2}}{\pi_{i2}}]E(Y_{i2} - \mu_{i2}|\tilde{Y}_i^2, X_i)|\tilde{Y}_i^2, X_i) \\ \vdots \\ E(\frac{R_{im}}{\pi_{im}}[Y_{im} - \mu_{im}]|\tilde{Y}_i^o, X_i) + E([1 - \frac{R_{im}}{\pi_{im}}]E(Y_{im} - \mu_{im}|\tilde{Y}_i^m, X_i)|\tilde{Y}_i^m, X_i) \end{bmatrix} \right) \\
=& E\left(\dot{\mu}_i \Sigma_i^{-1} \begin{bmatrix} E(Y_{i1} - \mu_{i1}|X_i) \\ \frac{\pi_{i2}^*}{\pi_{i2}} E(Y_{i2} - \mu_{i2}|\tilde{Y}_i^2, X_i) + (1 - \frac{\pi_{i2}^*}{\pi_{i2}}) E(Y_{i2} - \mu_{i2}|\tilde{Y}_i^2, X_i) \\ \vdots \\ \frac{\pi_{im}^*}{\pi_{im}} E(Y_{im} - \mu_{im}|\tilde{Y}_i^m, X_i) + (1 - \frac{\pi_{im}^*}{\pi_{im}}) E(Y_{im} - \mu_{im}|\tilde{Y}_i^m, X_i) \end{bmatrix} \right) \\
=& E\left(\dot{\mu}_i \Sigma_i^{-1} \begin{bmatrix} E(Y_{i1} - \mu_{i1}|X_i) \\ E(Y_{i2} - \mu_{i2}|\tilde{Y}_i^2, X_i) \\ \vdots \\ E(Y_{im} - \mu_{im}|\tilde{Y}_i^m, X_i) \end{bmatrix} \right) \\
=& E[\dot{\mu}_i \Sigma_i^{-1} (Y_i - \mu_i)]
\end{aligned}$$

where π_{im}^* is the the propensity score for the m-th observation given the observed covariates X_i and response history $\tilde{Y}_i^m$. The consistency follows from the 0 expectation when regression coefficients take the true value.

When the missing probability model is correct but the LCM condition does not hold,

$$
\begin{aligned}
E(U_i) =& E\left(\dot{\mu}_i\Sigma_i^{-1}\begin{bmatrix} E(Y_{i1}-\mu_{i1}|X_i) \\ E(\frac{R_{i2}}{\pi_{i2}}[Y_{i2}-\mu_{i2}]|\tilde{Y}_i^2, X_i) + E([1-\frac{R_{i2}}{\pi_{i2}}]\tilde{E}(Y_{i2}-\mu_{i2})|\tilde{Y}_i^2, X_i) \\ \vdots \\ E(\frac{R_{im}}{\pi_{im}}[Y_{im}-\mu_{im}]|\tilde{Y}_i^o, X_i) + E([1-\frac{R_{im}}{\pi_{im}}]\tilde{E}(Y_{im}-\mu_{im})|\tilde{Y}_i^m, X_i)\end{bmatrix}\right) \\
=& E\left(\dot{\mu}_i\Sigma_i^{-1}\begin{bmatrix} E(Y_{i1}-\mu_{i1}|X_i) \\ E(Y_{i2}-\mu_{i2}|\tilde{Y}_i^2, X_i) + 0 \\ \vdots \\ E(Y_{im}-\mu_{im}|\tilde{Y}_i^m, X_i) + 0\end{bmatrix}\right) \\
=& E\left(\dot{\mu}_i\Sigma_i^{-1}\begin{bmatrix} E(Y_{i1}-\mu_{i1}|X_i) \\ E(Y_{i2}-\mu_{i2}|\tilde{Y}_i^2, X_i) \\ \vdots \\ E(Y_{im}-\mu_{im}|\tilde{Y}_i^m, X_i)\end{bmatrix}\right) \\
=& E[\dot{\mu}_i\Sigma_i^{-1}(Y_i-\mu_i)]
\end{aligned}
$$

The consistency follows from the same reason as mentioned above.

2. Web Appendix B: Regularity conditions and proof of Theorem 1

We will show the asymptotic properties of the proposed estimator. To accommodate the possible dependence between X and T, we assume the following relationship as Rice (1986): $X_{ijk} = m_k(T_{ij}) + \varrho_{ijk}$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$, $1 \leqslant k \leqslant p$ where the $m_k(\cdot)$ are functions with bounded rth derivatives and the ϱ_{ijk} are independent random variables with mean 0 and are independent of the standardized responses. Let Λ_n be a $N(= n \times m)$ by p matrix whose lth column is $\varrho_l = (\varrho_{11l}, \cdots, \varrho_{1ml}, \cdots, \varrho_{nml})^T$. Recall that in Section 3, we denote $\Sigma_i = \Sigma_i(\mu_i(\theta))$, $\Delta_i = \Delta_i(\mu_i(\theta))$, $h_i = h_i(\mu_i(\theta), \gamma)$, $X = (X_1^T, \cdots, X_n^T)^T$ with $X_i = (X_{i1}, \cdots, X_{im})^T$, $M = (M_1^T, \cdots, M_n^T)^T$ with $M_i = (\pi_{i1}, \cdots, \pi_{im})^T$, $\Omega = diag\{\Omega_1, \cdots, \Omega_n\}$

with $\Omega_i = \Delta_i^T \Sigma_i^{-1} E\{\frac{\partial}{\partial \mu_i} h_i\}\Delta_i$, $P = M(M^T \Omega M)^{-1} M^T \Omega$, $X^* = (X_1^{*T}, \cdots, X_n^{*T})^T = (I - P)X$, $B_i = X_i^{*T} \Delta_i^T \Sigma_i^{-1} h_i - [\sum_{i=1}^n X_i^{*T} \Delta_i^T \Sigma_i^{-1} \frac{\partial}{\partial \gamma} h_i(\mu_i, \gamma)] \cdot [\frac{\partial}{\partial \gamma} G_\gamma(\gamma)]^{-1} G_{\gamma,i}(\gamma)$ and the notations $\Omega_{0,i}$ represent Ω_i evaluated at the true $\mu_{0,i}$ and γ_0. Notations $B_{0,i}$ and $X_{0,i}^*$ are defined in a similar fashion. Let $\|\cdot\|$ be the Euclidean norm. The regularity conditions as given as follows:

(C.1) The parameter vector γ_0 is an interior point of the parameter space Γ which is a compact set.

(C.2) $p_{ij}(\gamma) > c_1 > 0$ for all $\gamma \in \Gamma$, for some constant c_1.

(C.3) $\frac{1}{n} \frac{\partial}{\partial \gamma} G_\gamma(\gamma_0)$ and $\frac{1}{n} \sum_{i=1}^n G_{\gamma,i}(\gamma_0) G_{\gamma,i}^T(\gamma_0)$ converges to Σ_γ and V_γ respectively in probability for some positive-definite matrix Σ_γ and V_γ.

(C.4) The rth derivative of g_0 is bounded for some $r \geqslant 2$.

(C.5) The distinct values of $\{t_{ij}\}$ form a quasi-uniform sequence that grows dense on $[0, 1]$.

(C.6) There exists some ρ_0 such that the estimate of covariance parameters satisfies $\sqrt{n}(\hat{\rho} - \rho_0) = O_p(1)$.

(C.7) There exists positive constant c_2 such that $0 < c_2 \leqslant \nu(\cdot) < \infty$, $\nu(\cdot)$ and $\mu(\cdot)$ have bounded second derivatives and third derivatives respectively.

(C.8) For sufficiently large n, $k_n(M^T \Omega_0 M)$ is non-singular, and the eigenvalues of $(k_n/n) M^T \Omega_0 M$ are bounded away from zero and infinity in probability, where $\Omega_0 = diag\,\{\Omega_{0,1}, \cdots, \Omega_{0,n}\}$.

(C.9) $E\Lambda_n = 0$ and $sup_n \frac{1}{n} E\,\|\Lambda_n\|^2 < \infty$, and $\frac{1}{n} K_n \to K$, $\frac{1}{n} B_n \to B$ in probability for some positive definite matrix K and B, where $K_n = \sum_{i=1}^n X_{0,i}^{*T} \Omega_{0,i} X_{0,i}^*$ and $B_n = \sum_{i=1}^n B_{0,i} B_{0,i}^T$.

Lemma 1: *Assume that Conditions (C.4) and (C.5) hold, there exist $\alpha_0 \in R^{N_k}$ depending on g_0, and a constant C_4 depending only on l and C_0 such that*

$$\sup_{t \in [0,1]} |g_0(t) - \pi^T(t) \alpha_0| \leqslant C_4 k_n^{-r}.$$

The proof follows from Theorem 12.7 in Schumaker (2007).

Lemma 2: *Assume that Conditions (C.1)-(C.3) hold. Then*

$$\sqrt{n}(\hat{\gamma}-\gamma_0) \rightarrow N(0, \Sigma_\gamma^{-1} V_\gamma \Sigma_\gamma^{-1}).$$

The proof can be obtained easily using standard method and is omitted here.

Proof of Theorem 1:

Let

$$\xi(\beta,\alpha) = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} K_n^{1/2}(\beta-\beta_0) \\ k_n^{-1/2} H_n(\alpha-\alpha_0) + k_n^{1/2} H_n^{-1} M^T \Omega_0 X(\beta-\beta_0) \end{bmatrix},$$

$\hat{\xi} = (\hat{\xi}_1^T, \hat{\xi}_2^T)^T = \xi(\hat{\beta}, \hat{\alpha})$, where $H_n^2 = k_n M^T \Omega_0 M$. Then the doubly robust estimating equation can be expressed as

$$U_{\xi,n}(\mu(\xi)) = \sum_{i=1}^n D_i^T \Delta_i(\mu_i(\xi)) \Sigma_i^{-1}(\mu_i(\xi)) h_i(\mu_i(\xi), \hat{\gamma}). \tag{1}$$

Denote $\tilde{X}_i = K_n^{-1/2} X_i^{*T}$, $\tilde{M}_i = k_n^{1/2} H_n^{-1} \pi_i^T$, $R_{ni} = \pi_i \alpha_0 - g_0(t_i)$, and $\zeta_i = \tilde{X}_i \xi_1 + \tilde{M}_i \xi_2 + R_{ni}$, then $\eta_i(\theta) = D_i \theta = \eta_{0,i} + \zeta_i, i = 1, \cdots, n$, where $\eta_{0,i} = X_i \beta_0 + g_0(t_i)$. Let

$$N = \begin{bmatrix} K_n^{-1/2} & -K_n^{-1/2} X^T \Omega_0 M (M^T \Omega_0 M)^{-1} \\ 0 & k_n^{1/2} H_n^{-1} \end{bmatrix}.$$

Then (1) can be written as

$$\begin{aligned} \Psi(\mu(\xi), \hat{\gamma}) &= \begin{pmatrix} \Psi_1(\mu(\xi), \hat{\gamma}) \\ \Psi_2(\mu(\xi), \hat{\gamma}) \end{pmatrix} = N U_\xi(\mu(\xi)) \\ &= \begin{pmatrix} \sum_{i=1}^n K_n^{-1/2} X_i^{*T} \Delta_i^T(\mu_i(\xi)) \Sigma_i^{-1}(\mu_i(\xi)) h_i(\mu_i(\xi), \hat{\gamma}) \\ \sum_{i=1}^n k_n^{1/2} H_n^{-1} \pi_i^T \Delta_i^T(\mu_i(\xi)) \Sigma_i^{-1}(\mu_i(\xi)) h_i(\mu_i(\xi), \hat{\gamma}) \end{pmatrix} \qquad (2) \\ &= \sum_{i=1}^n \tilde{D}_i \Delta_i^T(\mu_i(\xi)) \Sigma_i^{-1}(\mu_i(\xi)) h_i(\mu_i(\xi), \hat{\gamma}), \end{aligned}$$

where $\tilde{D}_i = (X_i^* K_N^{-1/2}, \pi_i H_n^{-1} k_n^{1/2})^T$.

Combining (C.8) and (C.9), both equations (1) and (2) give the same root for ξ as our estimator. We denote

$$\Phi(\xi) = \begin{pmatrix} \Phi_1(\xi) \\ \Phi_2(\xi) \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \sum_{i=1}^{n} \{ \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} h_{0,i} - [\sum_{i=1}^{n} \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} \frac{\partial}{\partial \gamma} h_{0,i}] \Sigma_{\gamma}^{-1} G_{\gamma,i}(\gamma_0) \}.$$

The root $\tilde{\xi}$ of $\Phi(\xi)$,

$$\tilde{\xi} = \begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \end{pmatrix} = -\sum_{i=1}^{n} \{ \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} h_{0,i} - [\sum_{i=1}^{n} \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} \frac{\partial}{\partial \gamma} h_{0,i}] \Sigma_{\gamma}^{-1} G_{\gamma,i}(\gamma_0) \}.$$

is not an estimator. In the following, we will show that the difference between $\tilde{\xi}$ and $\hat{\xi}$ is small.

Let $a \in R^{p+N_k}$ satisfying $a^T a = 1$. Expand $a^T \Psi(\mu(\xi), \hat{\gamma})$ in a Taylor series and have

$$\begin{aligned} a^T \Psi(\mu(\xi), \hat{\gamma}) =& a^T \Psi(\mu(\eta_0 + \zeta), \hat{\gamma}) \\ =& \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_i^T (\mu_i(\eta_0 + \zeta)) \Sigma_i^{-1} (\mu_i(\eta_0 + \zeta)) h_i(\mu_i(\eta_0 + \zeta), \hat{\gamma}) \\ =& \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} h_i(\mu_{0,i}, \hat{\gamma}) + \sum_{i=1}^{n} a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} \frac{\partial}{\partial \mu_i} h_i(\mu_{0,i}, \hat{\gamma}) \Delta_{0,i} \zeta_i \\ &+ \sum_{i=1}^{n} \zeta_i^T \Delta_{0,i}^T \frac{\partial}{\partial \mu_i} (a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1}) h_i(\mu_{0,i}, \hat{\gamma}) + R_n^{**}(\mu^*, \hat{\gamma}) \\ =&: A_1 + A_2 + A_3 + A_4, \end{aligned}$$

where $R_n^{**}(\mu^*, \hat{\gamma}) = \sum_{i=1}^{n} R_{ni}^{**}(\mu_i^*, \hat{\gamma})$ and $R_{ni}^{**}(\mu_i^*, \hat{\gamma}) = \frac{1}{2} \zeta_i^T \Delta_i^T \frac{\partial^2}{\partial \mu_i \partial \mu_i^T} (a^T \tilde{D}_i \Delta_i(\mu_i) \Sigma_i^{-1}(\mu_i) \cdot h_i(\mu^*, \hat{\gamma}) \Delta_i \zeta_i$ evaluated at $\mu_i^* = \mu(\mu_{0,i} + \tau_i \zeta_i)$ for $i = 1, \cdots, n$ with $0 < \tau_i < 1$.

We first consider A_1 and expand it with respect to γ, then we have

$$\begin{aligned}
A_1 =& \sum_{i=1}^n a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} h_{0,i} \\
&+ \sum_{i=1}^n a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} \frac{\partial}{\partial \gamma} h_{0,i} (\hat{\gamma} - \gamma_0) \\
&+ \frac{1}{2} \sum_{i=1}^n (\hat{\gamma} - \gamma_0)^T \frac{\partial}{\partial \gamma \partial \gamma^T} [a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} h_i(\mu_{0,i}, \gamma^*)] (\hat{\gamma} - \gamma_0) \\
=:& A_{1,1} + A_{1,2} + A_{1,3},
\end{aligned}$$

where γ^* are the point on the line between $\hat{\gamma}$ and γ_0.

For $A_{1,2}$, by conditions (C.1)-(C.3), we can obtain $n^{1/2}(\hat{\gamma} - \gamma_0) = -\Sigma_\gamma^{-1}(n^{-1/2} G_{\gamma_0}) + o_p(1)$. Combining (C.7) and (C.8), we have

$$A_{1,2} = -[\sum_{i=1}^n a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} \frac{\partial}{\partial \gamma} h_{0,i}] \Sigma_\gamma^{-1} \sum_{i=1}^n G_{\gamma,i} + o_p(k_n^{1/2})$$

Note that $\|\hat{\gamma} - \gamma_0\| = O_p(n^{-1/2})$, then $A_{1,3} = o_p(k_n^{1/2})$. Combining all these results, we can get

$$A_1 = \sum_{i=1}^n a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} h_{0,i} - \{\sum_{i=1}^n a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} \frac{\partial}{\partial \gamma} h_{0,i}\} \Sigma_\gamma^{-1} G_{\gamma,i}] + o_p(k_n^{1/2}). \tag{3}$$

Now let us turn to consider A_2. Similarly, applying Taylor expansion with respect to γ, we have

$$\begin{aligned}
A_2 =& \sum_{i=1}^n a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} \frac{\partial}{\partial \mu_i} h_i(\mu_{0,i}, \hat{\gamma}) \Delta_{0,i} \zeta_i \\
=& \sum_{i=1}^n a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} \frac{\partial}{\partial \mu_i} h_{0,i} \Delta_{0,i} \zeta_i \\
&+ \sum_{i=1}^n \frac{\partial}{\partial \gamma} [a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} \frac{\partial}{\partial \mu_i} h_i(\mu_{0,i}, \gamma^*) \Delta_{0,i} \zeta_i] (\hat{\gamma} - \gamma_0) \\
=& \sum_{i=1}^n a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} \frac{\partial}{\partial \mu_i} h_{0,i} \Delta_{0,i} \zeta_i \\
&+ \sum_{i=1}^n \frac{\partial}{\partial \gamma} [a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} \frac{\partial}{\partial \mu_i} h_i(\mu_{0,i}, \gamma^*) \Delta_{0,i} \tilde{D}_i^T \xi] (\hat{\gamma} - \gamma_0) \\
&+ \sum_{i=1}^n \frac{\partial}{\partial \gamma} [a^T \tilde{D}_i \Delta_{0,i}^T \Sigma_{0,i}^{-1} \frac{\partial}{\partial \mu_i} h_i(\mu_{0,i}, \gamma^*) \Delta_{0,i} R_{ni}] (\hat{\gamma} - \gamma_0) \\
=:& A_{2,1} + A_{2,2} + A_{2,3},
\end{aligned}$$

where γ^* are the point on the line between $\hat{\gamma}$ and γ_0.

According to conditions (C.7)-(C.9), the result that $\|\hat{\gamma}-\gamma_0\| = O_p(n^{-1/2})$ and Lemma 1, it can be obtained that $A_{2,2} = O_p(n^{-1/2}k_n\|\xi\|) = o_p(\|\xi\|)$ and $A_{2,3} = O_p(k_n^{1/2-r})$. Then we have

$$A_2 = \sum_{i=1}^n a^T\tilde{D}_i\Delta_{0,i}^T\Sigma_{0,i}^{-1}\frac{\partial}{\partial\mu_i}h_{0,i}\Delta_{0,i}\zeta_i + o_p(\|\xi\|) + o_p(k_n^{1/2}), \tag{4}$$

By similar derivation, we can show

$$A_3 = \sum_{i=1}^n \zeta_i^T\Delta_{0,i}^T\frac{\partial}{\partial\mu_i}(a^T\tilde{D}_i\Delta_{0,i}^T\Sigma_{0,i}^{-1})h_{0,i} + o_p(\|\xi\|) + o_p(k_n^{1/2}), \tag{5}$$

and

$$A_4 = R_n^{**}(\mu^*,\gamma_0) + o_p(\|\xi\|) + o_p(k_n^{1/2}) = \sum_{i=1}^n R_{ni}^*(\mu_i^*,\gamma_0) + o_p(\|\xi\|) + o_p(k_n^{1/2}). \tag{6}$$

Combining (3) and (4)-(6), we have

$$\begin{aligned}
a^T\Psi(\mu(\xi),\hat{\gamma}) =& a^T\Psi(\mu(\eta_0+\zeta),\hat{\gamma})\\
=& \sum_{i=1}^n [a^T\tilde{D}_i\Delta_{0,i}^T\Sigma_{0,i}^{-1}h_{0,i} - \{\sum_{i=1}^n a^T\tilde{D}_i\Delta_{0,i}^T\Sigma_{0,i}^{-1}\frac{\partial}{\partial\gamma}h_{0,i}\}\Sigma_\gamma^{-1}G_{\gamma,i}]\\
&+ \sum_{i=1}^n a^T\tilde{D}_i\Delta_{0,i}^T\Sigma_{0,i}^{-1}\frac{\partial}{\partial\mu_i}h_{0,i}\Delta_{0,i}\zeta_i\\
&+ \sum_{i=1}^n \zeta_i^T\Delta_{0,i}^T\frac{\partial}{\partial\mu_i}(a^T\tilde{D}_i\Delta_{0,i}^T\Sigma_{0,i}^{-1})h_{0,i}\\
&+ R_n^{**}(\mu^*,\gamma_0) + o_p(\|\xi\|) + o_p(k_n^{1/2}).
\end{aligned}$$

Then

$$\begin{aligned}
a^T(\Psi(\mu(\xi),\hat{\gamma}) - \Phi(\xi)) =& \sum_{i=1}^n a^T\tilde{D}_i\Delta_{0,i}^T\Sigma_{0,i}^{-1}[\frac{\partial}{\partial\mu_i}h_{0,i} - E(\frac{\partial}{\partial\mu_i}h_{0,i})]\Delta_{0,i}\tilde{D}_i^T\xi\\
&+ \sum_{i=1}^n a^T\tilde{D}_i\Delta_{0,i}^T\Sigma_{0,i}^{-1}\frac{\partial}{\partial\mu_i}h_{0,i}\Delta_{0,i}R_{ni}\\
&+ \sum_{i=1}^n \zeta_i^T\Delta_{0,i}^T\frac{\partial}{\partial\mu_i}(a^T\tilde{D}_i\Delta_{0,i}^T\Sigma_{0,i}^{-1})h_{0,i}\\
&+ R_n^{**}(\mu^*,\gamma_0) + o_p(\|\xi\|) + o_p(k_n^{1/2}).
\end{aligned}$$

Then using the same argument as He et al. (2005), it can be shown that

$$\sup_{\|\xi\|\leqslant Lk_n^{1/2}} |\Psi(\mu(\xi),\hat{\gamma}) - \Psi(\xi)\| = O_p(k_n^{1/2}),$$

for sufficiently large constant L.

By direct calculation,

$$E\left\|\tilde{\xi}\right\|^2 = O(k_n).$$

Then we have

$$\sup_{\|\xi\|\leqslant Lk_n^{1/2}} \|\Psi(\xi)-\xi\| \leqslant \sup_{\|\xi\|\leqslant Lk_n^{1/2}} \|\Psi(\xi)-\Phi(\xi)\| + \|\tilde{\xi}\| = O_p(k_n^{1/2}),$$

which indicates that $\sup_{\|\xi\|\leqslant Lk_n^{1/2}} \|\Psi(\xi)-\xi\| \leqslant Lk_n^{1/2}$ in probability, for sufficiently large L. Thus, Brouwer's fixed point theorem guarantees there exists a zero $\hat{\xi}$ of $\Psi(\xi)$ with $\|\hat{\xi}\| = O_p(k_n^{1/2})$ and hence the optimal convergencerate of the estimator of the nonparametric function can be achieved. Applying the central limit theorem on $\tilde{\xi}_1$, the asymptotic normality of the estimator of $\hat{\beta}$ can be established similarly.

3. Web Appendix C: Simulations for binary outcome

Consider a binary partial linear model as

$$ln(\mu_{ij}/(1-\mu_{ij})) = X_{ij}\beta_0 + 0.5\cos(\pi T_{ij}),$$

where $\beta_0 = 0.5$, X_{ij} and T_{ij} are independently drawn from uniform distributions on $(-0.8, 0.8)$ and $(-0.5, 0.5)$ respectively, $R_i(\rho)$ is the correlation matrix of Y_i considered to be AR1 with correlation parameter $\rho = 0.6$. The correlated binary data are generated using the method proposed in Preisser et al. (2002). The sample size is also $n = 600$ with $m = 6$.

The values of the indicators R_{ij} are generated from a model similar to model (9) in the study for continuous response except that the parameter vector $(\gamma_0, \gamma_1, \gamma_2)^T$ is taken to be $(1.5, 1.0, -1.0)^T$, which yields about 25% missingness.

Similarly, we consider 5 scenarios as in the study for continuous outcome. The only difference is that in order to violate LCM assumption, we force an exchangeable working correlation, i.e. a wrong correlation matrix, in S3 for the binary outcome. In S5, the missing indicator R_{ij} is generated from the model $ln\frac{p_{ij}}{1-p_{ij}} = \gamma_0 + \gamma_1 Y_{i,j-1} + \gamma_2 X_{ij} + \gamma_3 Y_{ij}$ with

$(\gamma_0, \gamma_1, \gamma_2, \gamma_3)^T$ taken to be $(1.5, 1.0, -1.0, 0.5)^T$. Note that we are not saying that LCM holds under the simulated multivariate binary dataset. Instead, we are going to rely on the approximate truth of the assumption, as Qu et al. (2010) have done in their simulation part.

The simulation results based on 500 replications is presented in Table 1 . Again the proposed method shows its double robustness. When the LCM condition holds, it has a comparable performance with Qu's method; when the LCM assumption is violated but the dropout model is correctly specified, it outperforms Qu's method in terms of bias. Despite of its larger standard error, as is expected, the proposed doubly robust method still yields a slightly smaller MSE compared to Qu's method. S5 shows that when MAR is violated, estimates from the four methods all deviate from the truth.

[Table 1 about here.]

References

He, X., Fung, W. K., and Zhu, Z. (2005). Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association* **100,** 1176–1184.

Preisser, J. S., Lohman, K. K., and Rathouz, P. J. (2002). Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. *Statistics in medicine* **21,** 3035–3054.

Qu, A., Lindsay, B. G., and Lu, L. (2010). Highly efficient aggregate unbiased estimating functions approach for correlated data with missing at random. *Journal of the American Statistical Association* **105,** 194–204.

Rice, J. (1986). Convergence rates for partially splined models. *Statistics & Probability Letters* **4,** 203–208.

Schumaker, L. (2007). *Spline functions: basic theory.* Cambridge University Press.

Received XXX 2015.

Table 1
Simulation results for binary outcome

		$\beta_0 = 0.5$				
		BIAS	ESE	SE	MSE	IMSE
S 1	GEE-C	0.0641	0.0678	0.0676	0.0087	0.0190
	GEE-W	0.0184	0.0703	0.0700	0.0052	0.0123
	Qu	0.0167	0.0636	0.0645	0.0044	0.0106
	DR-PLM	0.0167	0.0693	0.0692	0.0051	0.0118
S 2	GEE-C	0.0641	0.0678	0.0676	0.0087	0.0190
	GEE-W	0.0622	0.0701	0.0695	0.0087	0.0121
	Qu	0.0167	0.0636	0.0645	0.0044	0.0106
	DR-PLM	0.0171	0.0679	0.0684	0.0050	0.0116
S 3	GEE-C	0.0705	0.0788	0.0773	0.0109	0.0246
	GEE-W	0.0217	0.0819	0.0806	0.0070	0.0152
	Qu	0.0389	0.0740	0.0757	0.0072	0.0131
	DR-PLM	0.0197	0.0802	0.0809	0.0069	0.0146
S 4	GEE-C	0.0705	0.0788	0.0773	0.0109	0.0246
	GEE-W	0.0696	0.0817	0.0800	0.0112	0.0151
	Qu	0.0389	0.0740	0.0757	0.0072	0.0131
	DR-PLM	0.0426	0.0795	0.0804	0.0083	0.0144
S 5	GEE-C	0.0907	0.0664	0.0670	0.0127	0.0295
	GEE-W	0.0473	0.0683	0.0688	0.0070	0.0144
	Qu	0.0423	0.0624	0.0642	0.0059	0.0122
	DR-PLM	0.0475	0.0676	0.0682	0.0069	0.0137

Note: SE: standard error; ESE: estimated standard error
from asymptotic theory; MSE: mean square error;
IMSE: integrated MSE; GEE-C: complete-case GEE;
GEE-W: IPW GEE; Qu: Qu's method; DR-PLM: the proposed method.