Web-based Supplementary Materials for "Improved Doubly Robust Estimation when Data are Monotonely Coarsened, with Application to Longitudinal Studies with Dropout"

Anastasios A. Tsiatis*, **Marie Davidian, and Weihua Cao,**

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203, U.S.A.

**email:* tsiatis@stat.ncsu.edu

Web Appendix A: Derivation of Approximate Standard Errors Via the Sandwich Method

Throughout, we use the notation defined in the main paper. We provide expressions required to calculate the asymptotic variances of the three estimators for β $(p \times 1)$ considered in the main paper: $\widehat{\beta}_{ipw}$, $\widehat{\beta}_{br*}$, and $\widehat{\beta}_{opt*}$. Let τ be the collection of unknown parameters involved in obtaining the estimators for β; in particular, $\tau = (\psi^T, \beta^T)^T$ for $\widehat{\beta}_{ipw}$, $\tau = (\psi^T, \xi_1^T, ..., \xi_M^T, \beta^T)^T$ for $\widehat{\beta}_{br^*}$, and $\tau = (\psi^T, \xi^T, \theta^T, \beta^T)^T$ for $\widehat{\beta}_{opt^*}$. The estimator for τ, $\widehat{\tau}$, in each case can be obtained by solving a set of M-estimating equations given by $\sum_{i=1}^n \rho_i(\tau) = 0$ (Stefanski and Boos, 2002), where the last p entries of $\rho_i(\tau)$ correspond to the estimating equation for β, and $\rho_i(\tau)$ is defined for each estimator below. Let $A_n = n^{-1}\sum_{i=1}^n A_i = n^{-1}\sum_{i=1}^n \partial/\partial\tau\{\rho_i(\tau)\}$, and $B_n = n^{-1}\sum_{i=1}^n \rho_i(\tau)\rho_i^T(\tau)$. Following standard theory, the asymptotic covariance matrix of $\widehat{\tau}$ can be approximated by the empirical sandwich matrix $V_n = n^{-1}A_n^{-1}B_n(A_n^{-1})^T$. Therefore, the asymptotic variances of the three estimators can be approximated by the lower, rightmost diagonal $(p \times p)$ submatrix of the corresponding matrix V_n. We present the form of $\rho_i(\tau)$ and A_i for each of the estimators, from which the form of V_n may be calculated. The desired diagonal submatrix of V_n may then be obtained numerically, with the required matrix inversion carried out by standard routines.

Throughout, we assume that $\lambda_r\{G_r(Z), \psi_r\}$, $r = 1, \ldots, M$, are logistic regression models, and $\psi = (\psi_1^T, ..., \psi_M^T)^T$ are estimated via separate ML fits for each $r = 1, \ldots, M$, where $\widetilde{X}_{i,r}$

is a row vector consisting of the covariates used in the modeling of $\lambda_r\{G_r(Z_i), \psi_r\}$, including a "1" for the intercept term. For $\widehat{\beta}_{ipw}$, $\rho_i(\tau)$ is given by

$$
\begin{aligned}
\rho_i(\tau) &= \begin{pmatrix} \sum_{r=1}^{M} \frac{dM_C\{r, G_r(Z_i), \psi\}}{K_r\{G_r(Z_i), \psi\}} \frac{K_{r-1}\{G_r(Z_i), \psi\}\lambda_{r\psi}\{G_r(Z_i), \psi\}}{\lambda_r\{G_r(Z_i), \psi\}} \\ \frac{I(C_i = \infty)m(Z_i, \beta)}{\pi(\infty, Z_i, \psi)} \end{pmatrix} \\
&= \begin{pmatrix} dM_C\{1, G_1(Z_i), \psi_1\}\widetilde{X}_{i,1}^T \\ \vdots \\ dM_C\{M, G_M(Z_i), \psi_M\}\widetilde{X}_{i,M}^T \\ \frac{I(C_i = \infty)m(Z_i, \beta)}{\pi(\infty, Z_i, \psi)} \end{pmatrix},
\end{aligned}
$$

and A_i is given by

$$
A_i = \begin{pmatrix}
D_{i,1} & 0 & \cdots & \cdots & \cdots & 0 \\
0 & \ddots & 0 & \cdots & \cdots & 0 \\
0 & 0 & D_{i,r} & 0 & \cdots & 0 \\
0 & \cdots & 0 & \ddots & 0 & 0 \\
0 & \cdots & \cdots & 0 & D_{i,M} & 0 \\
E_{i,1} & \cdots & E_{i,r} & \cdots & E_{i,M} & D_{i,\beta}
\end{pmatrix},
$$

where

$$
\begin{aligned}
D_{i,r} &= -I(C_i \geq r)\lambda_r\{G_r(Z_i), \psi_r\}[1 - \lambda_r\{G_r(Z_i), \psi_r\}]\widetilde{X}_{i,r}^T\widetilde{X}_{i,r}, \quad r = 1, \ldots, M, \\
E_{i,r} &= \frac{I(C_i = \infty)m(Z_i, \beta)}{\pi(\infty, Z_i, \psi)}\lambda_r\{G_r(Z_i), \psi_r\}\widetilde{X}_{i,r}, \quad r = 1, \ldots, M, \\
D_{i,\beta} &= \frac{I(C_i = \infty)}{\pi(\infty, Z_i, \psi)}m_\beta(Z_i, \beta),
\end{aligned}
$$

and $m_\beta(Z_i, \beta)$ is a column vector of partial derivatives of $m(Z_i, \beta)$ with respect to β.

We implemented $\widehat{\beta}_{br^*}$ as described in Bang and Robins (2005); i.e., we added as a covariate $\widehat{K}_r^{-1}\left\{G_r(Z),\widehat{\psi}_1,\ldots,\widehat{\psi}_r\right\}$ in the conditional mean functions $h_r^*\{G_r(Z),\xi_r\}$ corresponding to a generalized linear model with canonical link, where $\widehat{K}_r^{-1}\left\{G_r(Z),\widehat{\psi}_1,\ldots,\widehat{\psi}_r\right\}$ is an estimate for the true cumulative hazard $K_r^{-1}\{G_r(Z),\psi_1,\ldots,\psi_r\}$, $r=1,\ldots,M$. We write the new conditional mean function including additional covariate $\widehat{K}_r^{-1}\left\{G_r(Z),\widehat{\psi}_1,\ldots,\widehat{\psi}_r\right\}$ as $h_r^*\{G_r(Z),\psi_1,...,\psi_r,\xi_r,\beta\}$. For this estimator, $\rho_i(\tau)$ is given by

$$\rho_i(\tau)=\begin{pmatrix} dM_C\{1,G_1(Z_i),\psi_1\}\widetilde{X}_{i,1}^T \\ \vdots \\ dM_C\{M,G_M(Z_i),\psi_M\}\widetilde{X}_{i,M}^T \\ I(C_i>1)\left[h_2^*\{G_2(Z_i),\psi_1,\psi_2,\xi_2,\beta\}-h_1^*\{G_1(Z_i),\psi_1,\xi_1,\beta\}\right] \\ \times h_{1,\psi_1,\xi_1}^*\{G_1(Z_i),\psi_1,\xi_1,\beta\} \\ \vdots \\ I(C_i>M)\left[m(Z_i,\beta)-h_M^*\{G_M(Z_i),\psi,\xi_M,\beta\}\right] \\ \times h_{M,\psi,\xi_M}^*\{G_M(Z_i),\psi,\xi_M,\beta\} \\ h_1^*\{G_1(Z_i),\psi_1,\xi_1,\beta\} \end{pmatrix},$$

where $h_{r,\psi_1,\ldots,\psi_r,\xi_r}^*\{G_r(Z_i),\psi_1,...,\psi_r,\xi_r,\beta\}$ is the column vector of partial derivatives of $h_r^*\{G_r(Z_i),\psi_1,...,\psi_r,\xi_r,\beta\}$ with respect to $\psi_1,...,\psi_r,\xi_r$, $r=1,...,M$.

The matrix A_i is given by

$$A_i=\begin{pmatrix} A_{1i} \\ A_{2i} \\ A_{3i}\end{pmatrix},\quad A_{1i}=\begin{pmatrix} D_{i,1} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & \cdots & 0 \\ 0 & 0 & D_{i,r} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \ddots & 0 & 0 \\ 0 & \cdots & \cdots & 0 & D_{i,M} & 0 \end{pmatrix},$$

$$\text{and} \quad A_{3i} = \begin{pmatrix} F_{1,i} & 0 & \cdots & 0 & F_{2,i} & 0 & \cdots & 0 & F_{3,i} \end{pmatrix},$$

where

$$\begin{aligned}
D_{i,r} &= -I(C_i \geq r)\lambda_r \{G_r(Z_i), \psi_r\} [1 - \lambda_r \{G_r(Z_i), \psi_r\}] \widetilde{X}_{i,r}^T \widetilde{X}_{i,r}, \quad r = 1, \ldots, M, \\
F_{1,i} &= h^*_{1,\psi_1} \{G_1(Z_i), \psi_1, \xi_1, \beta\}, \quad F_{2,i} = h^*_{1,\xi_1} \{G_1(Z_i), \psi_1, \xi_1, \beta\}, \\
F_{3,i} &= h^*_{1,\beta} \{G_1(Z_i), \psi_1, \xi_1, \beta\};
\end{aligned}$$

i.e., $F_{1,i}, F_{2,i}, F_{3,i}$ are partial derivatives of $h^*_1 \{G_1(Z_i), \psi_1, \xi_1, \beta\}$ with respect to ψ_1, ξ_1, and β, respectively. The A_{2i} term involves the partial derivatives of the column vector

$$\rho_{2,i}(\tau) = \begin{pmatrix} I(C_i > 1) [h^*_2 \{G_2(Z_i), \psi_1, \psi_2, \xi_2, \beta\} - h^*_1 \{G_1(Z_i), \psi_1, \xi_1, \beta\}] \\ \times h^*_{1,\psi_1,\xi_1} \{G_1(Z_i), \psi_1, \xi_1, \beta\} \\ \vdots \\ I(C_i > M) [m(Z_i, \beta) - h^*_M \{G_M(Z_i), \psi, \xi_M, \beta\}] \\ \times h^*_{M,\psi,\xi_M} \{G_M(Z_i), \psi, \xi_M, \beta\} \end{pmatrix}$$

with respect to τ. Often in practice, it is cumbersome to obtain the analytical derivatives of $\rho_{2,i}(\tau)$ with respect to τ. In our implementation, we used numerical derivatives as an approximation to the analytical derivatives. For example, to calculate the derivative of $\rho_{2,i}(\tau)$ with respect to the kth element of τ, we used a one-sided numerical approximation of the form $\{\rho_{2,i}(\tau + \epsilon 1_k) - \rho_{2,i}(\tau)\} / \epsilon$ for small enough $\epsilon > 0$, where 1_k is a column vector with 1 on the kth entry and all other entries 0.

For $\widehat{\beta}_{opt^*}$, $\rho_i(\tau)$ is given by

$$\rho_i(\tau) = \begin{pmatrix} dM_C\left\{1, G_1(Z_i), \psi_1\right\} \widetilde{X}_{i,1}^T \\ \vdots \\ dM_C\left\{M, G_M(Z_i), \psi_M\right\} \widetilde{X}_{i,M}^T \\ \sum_{r=1}^M I(C_i > r)\widetilde{q}_r\left\{G_r(Z_i), \widetilde{\xi}, \psi\right\}\left[\widetilde{h}_{r+1}\left\{G_{r+1}(Z_i), \widetilde{\xi}, \psi\right\} - \widetilde{h}_r\left\{G_r(Z_i), \widetilde{\xi}, \psi\right\}\right] \\ \dfrac{I(C_i = \infty)m(Z_i, \beta)}{\pi(\infty, Z_i, \psi)} + \displaystyle\sum_{r=1}^M \dfrac{dM_c\left\{r, G_r(Z_i), \psi\right\}}{K_r\left\{G_r(Z_i), \psi\right\}} h_r\left\{G_r(Z_i), \xi\right\} \end{pmatrix},$$

where

$$\begin{aligned}
\widetilde{h}_r\{G_r(Z_i), \widetilde{\xi}\} &= h_r\left\{G_r(Z_i), \xi\right\} - \theta^T \frac{K_{r-1}\left\{G_r(Z_i), \psi\right\} \lambda_{r\psi}\{G_r(Z_i), \psi\}}{\lambda_r\left\{G_r(Z_i), \psi\right\}}, \\
\widetilde{q}_r\left\{G_r(Z_i), \widetilde{\xi}, \psi\right\} &= -[K_r\left\{G_r(Z_i), \psi\right\}]^{-1} \sum_{j=1}^r \frac{\lambda_j\left\{G_j(Z_i), \psi\right\}}{K_j\left\{G_j(Z_i), \psi\right\}} \begin{pmatrix} \widetilde{h}_{j\xi}\left\{G_j(Z_i), \widetilde{\xi}, \psi\right\} \\ \widetilde{h}_{j\theta}\left\{G_j(Z_i), \widetilde{\xi}, \psi\right\} \end{pmatrix}, \\
\widetilde{h}_{j\theta}\left\{G_j(Z_i), \widetilde{\xi}, \psi\right\} &= -K_{j-1}\left\{G_j(Z_i), \psi\right\} \lambda_{j\psi}\left\{G_j(Z_i), \psi\right\} / \lambda_j\left\{G_j(Z_i), \psi\right\}, \\
\widetilde{h}_{j\xi}\left\{G_j(Z_i), \widetilde{\xi}, \psi\right\} &= h_{j\xi}\left\{G_j(Z_i), \xi, \psi\right\}.
\end{aligned}$$

The matrix A_i is given by

$$A_i = \begin{pmatrix} A_{1i} \\ A_{2i} \end{pmatrix}, \quad A_{1i} = \begin{pmatrix} D_{i,1} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & \cdots & 0 \\ 0 & 0 & D_{i,r} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \ddots & 0 & 0 \\ 0 & \cdots & \cdots & 0 & D_{i,M} & 0 \end{pmatrix},$$

and

$$A_{2i} = \begin{pmatrix} \partial / \partial\tau\left\{\rho_{3,i}(\tau)\right\} \end{pmatrix},$$

where

$$D_{i,r} = -I(C_i \geq r)\lambda_r\left\{G_r(Z_i), \psi_r\right\}\left[1 - \lambda_r\left\{G_r(Z_i), \psi_r\right\}\right]\widetilde{X}_{i,r}^T\widetilde{X}_{i,r}, \quad r = 1, \ldots, M,$$

$$\rho_{3,i}(\tau) = \left(\begin{array}{c} \sum_{r=1}^{M} I(C_i > r)\widetilde{q}_r\left\{G_r(Z_i), \widetilde{\xi}, \psi\right\}\left[\widetilde{h}_{r+1}\left\{G_{r+1}(Z_i), \widetilde{\xi}, \psi\right\} - \widetilde{h}_r\left\{G_r(Z_i), \widetilde{\xi}, \psi\right\}\right] \\ \dfrac{I(C_i = \infty)m(Z_i, \beta)}{\pi(\infty, Z_i, \psi)} + \displaystyle\sum_{r=1}^{M}\dfrac{dM_c\left\{r, G_r(Z_i), \psi\right\}}{K_r\left\{G_r(Z_i), \psi\right\}}h_r\left\{G_r(Z_i), \xi\right\} \end{array}\right).$$

Analogous to the strategy for $\widehat{\beta}_{br^*}$, in our implementation, we used numerical derivatives as an approximation to the analytical derivatives of $\rho_{3,i}(\tau)$ with respect to τ.

Web Appendix B: Derivation of Conditional Expectations Implied by Assumed Mixed Models in Section 5

We derive the required conditional expectations $E(Y|Y_1, \ldots, Y_j, \widetilde{X})$ for $j = 1, \ldots, 4$ implied by model (17) in Section 5 of the main paper. The random vector $\Psi = (\alpha_0, \alpha_1, e_1, e_2, e_3, e_4)^T$ has multivariate normal distribution with mean μ and variance Σ, where

$$\mu = (\mu_{\alpha 0}, \mu_{\alpha 1}, 0_{1\times 4})^T, \quad \Sigma = \begin{pmatrix} \Sigma_\alpha & 0_{2\times 4} \\ 0_{4\times 2} & \sigma_e^2 I_4 \end{pmatrix},$$

$0_{a\times b}$ is a zero matrix with dimension $(a \times b)$, and I_a is an $(a \times a)$ identity matrix. Therefore, the distribution of $(\alpha_0, \alpha_1, Y_1, Y_2, Y_3, Y_4)^T$, conditional on $\widetilde{X}$, follows multivariate normal distribution with mean $\widetilde{\mu} = A\mu + c$ and variance $\widetilde{\Sigma} = A\Sigma A^T$, where

$$A = I_6 + \begin{pmatrix} 0_{2\times 2} & 0_{2\times 4} \\ A_{21} & 0_{4\times 4} \end{pmatrix}, \quad A_{21} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ t_1 & t_2 & t_3 & t_4 \end{pmatrix}^T, \quad \text{and} \quad c = \gamma^T\widetilde{X}\begin{pmatrix} 0_{2\times 1} \\ 1_{4\times 1} \end{pmatrix}.$$

Hence, the conditional mean is given by

$$\begin{aligned} E(Y|Y_1, Y_2, Y_3, Y_4, \widetilde{X}) &= E(Y_5|Y_1, Y_2, Y_3, Y_4, \widetilde{X}) \\ &= \gamma^T X + E(\alpha_0|Y_1, Y_2, Y_3, Y_4, \widetilde{X}) + t_5 E(\alpha_1|Y_1, Y_2, Y_3, Y_4, \widetilde{X}). \end{aligned}$$

To calculate the conditional mean $E(\alpha_k|Y_1, Y_2, Y_3, Y_4, \widetilde{X})$, $k = 0, 1$, we use the following property of multivariate normal distribution. Suppose $(X_1^T, X_2^T)^T$ follows a $N(\upsilon, \Omega)$ distribution. If υ and Ω are partitioned correspondingly as follows:

$$\upsilon = \begin{pmatrix} \upsilon_1 \\ \upsilon_2 \end{pmatrix} \quad \text{and} \quad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix},$$

then $(X_1|X_2 = a) \sim N(\bar{\upsilon}, \overline{\Omega})$, where $\bar{\upsilon} = \upsilon_1 + \Omega_{12}\Omega_{22}^{-1}(a - \upsilon_2)$. Straightforward application of the above property yields

$$E\left\{(\alpha_0, \alpha_1)^T|Y_1, ..., Y_4, \widetilde{X}\right\} = \widetilde{\mu}_{1:2} + \widetilde{\Sigma}_{1:2,3:6}\widetilde{\Sigma}_{3:6,3:6}^{-1}\left\{(Y_1, Y_2, Y_3, Y_4)^T - \widetilde{\mu}_{3:6}\right\},$$

where $\widetilde{\mu}_{a:b}$ is a column vector consisting of ath to bth entries of $\widetilde{\mu}$, and $\widetilde{\Sigma}_{a:b,m:n}$ is the submatrix of $\widetilde{\Sigma}$ with rows a to b and columns m to n. Therefore the conditional expectation is

$$E(Y|Y_1, Y_2, Y_3, Y_4, \widetilde{X}) = \gamma^T\widetilde{X} + (1, t_5)\left[\widetilde{\mu}_{1:2} + \widetilde{\Sigma}_{1:2,3:6}\widetilde{\Sigma}_{3:6,3:6}^{-1}\left\{(Y_1, Y_2, Y_3, Y_4)^T - \widetilde{\mu}_{3:6}\right\}\right].$$

Similarly,

$$\begin{aligned}
E(Y|Y_1, Y_2, Y_3, \widetilde{X}) &= \gamma^T\widetilde{X} + (1, t_5)\left[\widetilde{\mu}_{1:2} + \widetilde{\Sigma}_{1:2,3:5}\widetilde{\Sigma}_{3:5,3:5}^{-1}\left\{(Y_1, Y_2, Y_3)^T - \widetilde{\mu}_{3:5}\right\}\right], \\
E(Y|Y_1, Y_2, \widetilde{X}) &= \gamma^T\widetilde{X} + (1, t_5)\left[\widetilde{\mu}_{1:2} + \widetilde{\Sigma}_{1:2,3:4}\widetilde{\Sigma}_{3:4,3:4}^{-1}\left\{(Y_1, Y_2)^T - \widetilde{\mu}_{3:4}\right\}\right], \\
E(Y|Y_1, \widetilde{X}) &= \gamma^T\widetilde{X} + (1, t_5)\left\{\widetilde{\mu}_{1:2} + \widetilde{\Sigma}_{1:2,3:3}\widetilde{\Sigma}_{3:3,3:3}^{-1}(Y_1 - \widetilde{\mu}_{3:3})\right\}.
\end{aligned}$$

Next, we provide the derivation of the conditional expectations

$$E(Y|Y_1, \ldots, Y_j, \widetilde{X}, \text{dis}_1, \text{dis}_2, \text{dis}_3, \text{dis}_4)$$

for $j = 1, \ldots, 4$ implied by assumed linear mixed model used in the second, general coarsened data analysis in Section 5 of the main paper; i.e., we assumed that, for $r = 1, ..., 5$, the data follow the linear mixed model

$$Y_{ir} = \alpha_{0i} + \alpha_{1i}t_{ir} + \gamma^T\widetilde{X}_i + \phi_1 I(r \geq 3)\text{dis}_{i2} + \phi_2 I(r = 5)\text{dis}_{i4} + e_{ir},$$

where the random effects and within-subject deviations are normal as above, and now $\widetilde{X} =$ (weight,karnof,symp).

Following the same logic as above, the distribution of $(\alpha_0, \alpha_1, Y_1, Y_2, Y_3, Y_4)^T$, conditional on $(\widetilde{X}, \mathrm{dis}_1, \mathrm{dis}_2, \mathrm{dis}_3, \mathrm{dis}_4)$, follows multivariate normal distribution with mean $\widetilde{\mu}^* = A\mu + \widetilde{c}$ and variance $\widetilde{\widetilde{\Sigma}} = A\Sigma A^T$, where $A, \mu, \Sigma, \widetilde{\Sigma}$ are the same as above, and

$$\widetilde{c} = \left(0_{1\times 2}, \gamma^T \widetilde{X}, \gamma^T \widetilde{X}, \gamma^T \widetilde{X} + \phi_1 \mathrm{dis}_2, \gamma^T \widetilde{X} + \phi_1 \mathrm{dis}_2\right)^T.$$

The conditional expectations are given as follows:

$$\begin{aligned}
&E(Y|Y_1, Y_2, Y_3, Y_4, \widetilde{X}, \mathrm{dis}_1, \mathrm{dis}_2, \mathrm{dis}_3, \mathrm{dis}_4)\\
&\quad = \gamma^T \widetilde{X} + \phi_1 \mathrm{dis}_2 + \phi_2 \mathrm{dis}_4 + (1, t_5)\left[\widetilde{\mu}^*_{1:2} + \widetilde{\widetilde{\Sigma}}_{1:2,3:6} \widetilde{\widetilde{\Sigma}}^{-1}_{3:6,3:6}\left\{(Y_1, Y_2, Y_3, Y_4)^T - \widetilde{\mu}^*_{3:6}\right\}\right],\\
&E(Y|Y_1, Y_2, Y_3, \widetilde{X}, \mathrm{dis}_1, \mathrm{dis}_2, \mathrm{dis}_3, \mathrm{dis}_4)\\
&\quad = \gamma^T \widetilde{X} + \phi_1 \mathrm{dis}_2 + \phi_2 \mathrm{dis}_4 + (1, t_5)\left[\widetilde{\mu}^*_{1:2} + \widetilde{\widetilde{\Sigma}}_{1:2,3:5} \widetilde{\widetilde{\Sigma}}^{-1}_{3:5,3:5}\left\{(Y_1, Y_2, Y_3)^T - \widetilde{\mu}^*_{3:5}\right\}\right],\\
&E(Y|Y_1, Y_2, \widetilde{X}, \mathrm{dis}_1, \mathrm{dis}_2, \mathrm{dis}_3, \mathrm{dis}_4)\\
&\quad = \gamma^T \widetilde{X} + \phi_1 \mathrm{dis}_2 + \phi_2 \mathrm{dis}_4 + (1, t_5)\left[\widetilde{\mu}^*_{1:2} + \widetilde{\widetilde{\Sigma}}_{1:2,3:4} \widetilde{\widetilde{\Sigma}}^{-1}_{3:4,3:4}\left\{(Y_1, Y_2)^T - \widetilde{\mu}^*_{3:4}\right\}\right],\\
&E(Y|Y_1, \widetilde{X}, \mathrm{dis}_1, \mathrm{dis}_2, \mathrm{dis}_3, \mathrm{dis}_4)\\
&\quad = \gamma^T \widetilde{X} + \phi_1 \mathrm{dis}_2 + \phi_2 \mathrm{dis}_4 + (1, t_5)\left\{\widetilde{\mu}^*_{1:2} + \widetilde{\widetilde{\Sigma}}_{1:2,3:3} \widetilde{\widetilde{\Sigma}}^{-1}_{3:3,3:3}\left(Y_1 - \widetilde{\mu}^*_{3:3}\right)\right\}.
\end{aligned}$$

Web Appendix C: Derivation of Conditional Expectations Implied by Assumed Mixed Model in Section 6

We derive the required conditional expectations $E(Y|\overline{L}_j)$ for $j = 1, 2$ implied by the model used in Section 6 of the main paper. The model implies that, in truth,

$$E\left(Y|\overline{L}_2\right) = E\left\{E(Y|\overline{L}_2, \alpha_0, \alpha_1)|\overline{L}_2\right\} = \gamma^T X + \mu_1(X, Y_1, Y_2) + t_3 \mu_2(X, Y_1, Y_2),$$

where $\mu_1(X, Y_1, Y_2) = E\left(\alpha_0|X, Y_1, Y_2\right)$, and $\mu_2(X, Y_1, Y_2) = E\left(\alpha_1|X, Y_1, Y_2\right)$.

Thus, we need to calculate the conditional distribution of α_0, α_1 given X, Y_1, Y_2. The joint density of $(\alpha_0, \alpha_1, X, Y_1, Y_2)^T$ is given by

$$f(\alpha_0, \alpha_1, X, Y_1, Y_2) = f(Y_2|\alpha_0, \alpha_1, X, Y_1) f(Y_1|\alpha_0, \alpha_1, X) f(X) f(\alpha_0, \alpha_1)$$

Therefore,

$$\begin{aligned} f(\alpha_0, \alpha_1|X, Y_1, Y_2) &= \frac{f(\alpha_0, \alpha_1, X, Y_1, Y_2)}{\int f(\alpha_0, \alpha_1, X, Y_1, Y_2) d\alpha_0 d\alpha_1} \\ &= \frac{f(Y_2|\alpha_0, \alpha_1, X, Y_1) f(Y_1|\alpha_0, \alpha_1, X) f(\alpha_0, \alpha_1)}{\int f(Y_2|\alpha_0, \alpha_1, X, Y_1) f(Y_1|\alpha_0, \alpha_1, X) f(\alpha_0, \alpha_1) d\alpha_0 d\alpha_1}. \end{aligned}$$

As a consequence,

$$f(\alpha_0, \alpha_1|X, Y_1, Y_2) \propto f(Y_2|\alpha_0, \alpha_1, X, Y_1) f(Y_1|\alpha_0, \alpha_1, X) f(\alpha_0, \alpha_1).$$

After some algebra, it can be shown that, if we let $a = \sigma_{22}/(\sigma_{11}\sigma_{22} - \sigma_{12}^2)$, $b = -\sigma_{12}/(\sigma_{11}\sigma_{22} - \sigma_{12}^2)$, $c = \sigma_{11}/(\sigma_{11}\sigma_{22} - \sigma_{12}^2)$, $g_1(X, Y_1, Y_2) = a\mu_{\alpha_0} + b\mu_{\alpha_1} + (Y_2 + Y_1 - 2\gamma^T X)/\sigma_e^2$, and $g_2(X, Y_1, Y_2) = b\mu_{\alpha_0} + c\mu_{\alpha_1} + (Y_2 - \gamma^T X)/\sigma_e^2$, then

$$\begin{aligned} \mu_2(X, Y_1, Y_2) &= E(\alpha_1|Z, Y_1, Y_2) = \frac{g_1(X, Y_1, Y_2)\,(b + 1/\sigma_e^2) - g_2(X, Y_1, Y_2)\,(a + 2/\sigma_e^2)}{(b + 1/\sigma_e^2)^2 - (c + 1/\sigma_e^2)\,(a + 2/\sigma_e^2)}, \\ \mu_1(X, Y_1, Y_2) &= E(\alpha_0|Z, Y_1, Y_2) = \frac{g_2(X, Y_1, Y_2) - \mu_2(X, Y_1, Y_2)\,(c + 1/\sigma_e^2)}{b + 1/\sigma_e^2}. \end{aligned}$$

Similarly, we have

$$E\left(Y|\overline{L}_1\right) = E\left\{E(Y|\overline{L}_1, \alpha_0, \alpha_1)|\overline{L}_1\right\} = \gamma^T X + \mu_3(X, Y_1) + t_3\mu_4(X, Y_1),$$

where $\mu_3(X, Y_1) = E\left(\alpha_0|X, Y_1\right)$, and $\mu_4(X, Y_1) = E\left(\alpha_1|X, Y_1\right)$. Letting $d = b\mu_{\alpha_0} + c\mu_{\alpha_1}$, $g_3(X, Y_1) = a\mu_{\alpha_0} + b\mu_{\alpha_1} + (Y_1 - \gamma^T X)/\sigma_e^2$, we have

$$\begin{aligned} \mu_3(X, Y_1) &= E(\alpha_0|X, Y_1) = \frac{g_3(X, Y_1)\cdot c - d\cdot b}{\left(a + 1/\sigma_e^2\right) c - b^2}, \\ \mu_4(X, Y_1) &= E(\alpha_1|X, Y_1) = \frac{d - \mu_3(X, Y_1)\cdot b}{c}. \end{aligned}$$