Supporting Information for "Improving Estimation Efficiency for Regression with MNAR Covariates" by Che, Han, and Lawless

Proof of Theorem 1

Proof. The estimating equation is

$$\sum_{i=1}^{n} \begin{pmatrix} R_i \boldsymbol{U}(Y_i, \boldsymbol{X}_i, \boldsymbol{Z}_i; \boldsymbol{\beta}) \\ \boldsymbol{h}(Y_i, \boldsymbol{Z}_i, R_i; \boldsymbol{\beta}, \boldsymbol{\theta}) \end{pmatrix} = \mathbf{0},$$

so by Qin and Lawless (1994), the asymptotic covariance matrix of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$ can be written as

$$\text{ACov}\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\theta}} \end{pmatrix} = \left[\begin{pmatrix} ER\boldsymbol{U}_{\boldsymbol{\beta}} & E\boldsymbol{h}_{\boldsymbol{\beta}} \\ \mathbf{0} & E\boldsymbol{h}_{\boldsymbol{\theta}} \end{pmatrix}^T \begin{pmatrix} ER\boldsymbol{U}\boldsymbol{U}^T & ER\boldsymbol{U}\boldsymbol{h}^T \\ E\boldsymbol{h}\boldsymbol{U}^T & E\boldsymbol{h}\boldsymbol{h}^T \end{pmatrix}^{-1} \begin{pmatrix} ER\boldsymbol{U}_{\boldsymbol{\beta}} & \mathbf{0} \\ E\boldsymbol{h}_{\boldsymbol{\beta}} & E\boldsymbol{h}_{\boldsymbol{\theta}} \end{pmatrix} \right]^{-1}$$

$$= \left[\begin{pmatrix} ER\boldsymbol{U}_{\boldsymbol{\beta}} & E\boldsymbol{h}_{\boldsymbol{\beta}} \\ \mathbf{0} & E\boldsymbol{h}_{\boldsymbol{\theta}} \end{pmatrix}^T \begin{pmatrix} U^{11} & U^{12} \\ U^{21} & U^{22} \end{pmatrix} \begin{pmatrix} ER\boldsymbol{U}_{\boldsymbol{\beta}} & \mathbf{0} \\ E\boldsymbol{h}_{\boldsymbol{\beta}} & E\boldsymbol{h}_{\boldsymbol{\theta}} \end{pmatrix} \right]^{-1}$$

$$= \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}^{-1},$$

where

$$U^{11} = (ER\boldsymbol{U}\boldsymbol{U}^T)^{-1} + (ER\boldsymbol{U}\boldsymbol{U}^T)^{-1} ER\boldsymbol{U}\boldsymbol{h}^T \left\{ E\boldsymbol{h}\boldsymbol{h}^T - ER\boldsymbol{h}\boldsymbol{U}^T (ER\boldsymbol{U}\boldsymbol{U}^T)^{-1} ER\boldsymbol{U}\boldsymbol{h}^T \right\}^{-1} ER\boldsymbol{h}\boldsymbol{U}^T (ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}$$

$$U^{12} = (ER\boldsymbol{U}\boldsymbol{U}^T)^{-1} ER\boldsymbol{U}\boldsymbol{h}^T \left\{ E\boldsymbol{h}\boldsymbol{h}^T - ER\boldsymbol{h}\boldsymbol{U}^T (ER\boldsymbol{U}\boldsymbol{U}^T)^{-1} ER\boldsymbol{U}\boldsymbol{h}^T \right\}^{-1}$$

$$U^{21} = U^{12T}$$

$$U^{22} = \left\{ E\boldsymbol{h}\boldsymbol{h}^T - ER\boldsymbol{h}\boldsymbol{U}^T (ER\boldsymbol{U}\boldsymbol{U}^T)^{-1} ER\boldsymbol{U}\boldsymbol{h}^T \right\}^{-1}$$

so then

$$
\begin{aligned}
I_{11} &= ER\boldsymbol{U}_{\boldsymbol{\beta}}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}_{\boldsymbol{\beta}} \\
&+ ER\boldsymbol{U}_{\boldsymbol{\beta}}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T\left\{E\boldsymbol{h}\boldsymbol{h}^T - ER\boldsymbol{h}\boldsymbol{U}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T\right\}^{-1} \\
&\cdot ER\boldsymbol{h}\boldsymbol{U}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}_{\boldsymbol{\beta}} \\
&+ E\boldsymbol{h}_{\boldsymbol{\beta}}^T\left\{E\boldsymbol{h}\boldsymbol{h}^T - ER\boldsymbol{h}\boldsymbol{U}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T\right\}^{-1}ER\boldsymbol{h}\boldsymbol{U}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}_{\boldsymbol{\beta}} \\
&+ ER\boldsymbol{U}_{\boldsymbol{\beta}}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T\left\{E\boldsymbol{h}\boldsymbol{h}^T - ER\boldsymbol{h}\boldsymbol{U}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T\right\}^{-1}E\boldsymbol{h}_{\boldsymbol{\beta}} \\
&+ E\boldsymbol{h}_{\boldsymbol{\beta}}^T\left\{E\boldsymbol{h}\boldsymbol{h}^T - ER\boldsymbol{h}\boldsymbol{U}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T\right\}^{-1}E\boldsymbol{h}_{\boldsymbol{\beta}} \\
&= ER\boldsymbol{U}_{\boldsymbol{\beta}}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}_{\boldsymbol{\beta}} \\
&+ \left(ER\boldsymbol{U}_{\boldsymbol{\beta}}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T + E\boldsymbol{h}_{\boldsymbol{\beta}}^T\right)\left\{E\boldsymbol{h}\boldsymbol{h}^T - ER\boldsymbol{h}\boldsymbol{U}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T\right\}^{-1} \\
&\cdot \left(ER\boldsymbol{h}\boldsymbol{U}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}_{\boldsymbol{\beta}} + E\boldsymbol{h}_{\boldsymbol{\beta}}\right) \\
&:= ER\boldsymbol{U}_{\boldsymbol{\beta}}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}_{\boldsymbol{\beta}} + D_1C^{-1}D_1^T
\end{aligned}
$$

where $D_1 := ER\boldsymbol{U}_{\boldsymbol{\beta}}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T + E\boldsymbol{h}_{\boldsymbol{\beta}}^T$, and $C := E\boldsymbol{h}\boldsymbol{h}^T - ER\boldsymbol{h}\boldsymbol{U}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T$.

$$
\begin{aligned}
I_{12} &= ER\boldsymbol{U}_{\boldsymbol{\beta}}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T\left\{E\boldsymbol{h}\boldsymbol{h}^T - ER\boldsymbol{h}\boldsymbol{U}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T\right\}^{-1}E\boldsymbol{h}_{\boldsymbol{\theta}} \\
&+ E\boldsymbol{h}_{\boldsymbol{\beta}}^T\left\{E\boldsymbol{h}\boldsymbol{h}^T - ER\boldsymbol{h}\boldsymbol{U}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T\right\}^{-1}E\boldsymbol{h}_{\boldsymbol{\theta}} \\
&= D_1C^{-1}E\boldsymbol{h}_{\boldsymbol{\theta}} \\
I_{21} &= I_{12}^T \\
I_{22} &= E\boldsymbol{h}_{\boldsymbol{\theta}}^T\left\{E\boldsymbol{h}\boldsymbol{h}^T - ER\boldsymbol{h}\boldsymbol{U}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}\boldsymbol{h}^T\right\}^{-1}E\boldsymbol{h}_{\boldsymbol{\theta}} \\
&= E\boldsymbol{h}_{\boldsymbol{\theta}}^TC^{-1}E\boldsymbol{h}_{\boldsymbol{\theta}}
\end{aligned}
$$

Note that

$$\text{ACov}(\hat{\boldsymbol{\beta}}) = (I_{11} - I_{12}I_{22}^{-1}I_{21})^{-1}$$

and

$$
\begin{aligned}
I_{11} - I_{12}I_{22}^{-1}I_{21} &= ER\boldsymbol{U}_{\boldsymbol{\beta}}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}_{\boldsymbol{\beta}} + D_1C^{-1}D_1^T - D_1C^{-1}E\boldsymbol{h}_{\boldsymbol{\theta}}I_{22}^{-1}E\boldsymbol{h}_{\boldsymbol{\theta}}^TC^{-1}D_1^T \\
&= ER\boldsymbol{U}_{\boldsymbol{\beta}}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}_{\boldsymbol{\beta}} + D_1C^{-1}\left\{C - E\boldsymbol{h}_{\boldsymbol{\theta}}\left(E\boldsymbol{h}_{\boldsymbol{\theta}}^TC^{-1}E\boldsymbol{h}_{\boldsymbol{\theta}}\right)^{-1}E\boldsymbol{h}_{\boldsymbol{\theta}}^T\right\}C^{-1}D_1^T \\
&= ER\boldsymbol{U}_{\boldsymbol{\beta}}^T(ER\boldsymbol{U}\boldsymbol{U}^T)^{-1}ER\boldsymbol{U}_{\boldsymbol{\beta}} + \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^T.
\end{aligned}
$$

The desired result then follows. □

Lemma 1 and Proof

Lemma 1. *For a symmetric, positive definite matrix $A_{m\times m}$ and a full rank matrix $G_{m\times p}$ with $p \le m$,*

$$A - G(G^TA^{-1}G)^{-1}G^T$$

is positive semi-definite.

Proof. $\text{rank}(G) = p$, so it has singular value decomposition as

$$G = O_{m\times m} \begin{bmatrix} D_{p\times p} \\ \mathbf{0} \end{bmatrix} N_{p\times p}^T$$

where O, N are orthorgnal and D is diagonal. Then

$$\begin{aligned} G^T A^{-1} G =& N[D\ \mathbf{0}]O^T A^{-1} O \begin{bmatrix} D \\ \mathbf{0} \end{bmatrix} N^T =: N[D\ \mathbf{0}]Q \begin{bmatrix} D \\ \mathbf{0} \end{bmatrix} N^T \\ =& NDQ_1DN^T \end{aligned}$$

where $Q_{m\times m} = O^T A^{-1} O$, Q_1 is the first $q \times q$ diagonal block of Q, both invertible. In other words

$$O^T A^{-1} O = Q = \begin{bmatrix} Q_1 & Q_2 \\ Q_2^T & Q_3 \end{bmatrix}$$

O, N are orthogonal and D is diagonal. Then

$$\begin{aligned} A - G(G^T A^{-1} G)^{-1} G^T =& A - O \begin{bmatrix} D \\ \mathbf{0} \end{bmatrix} N^T N^{-T} (NDQ_1DN^T)^{-1} N^{-1} N[D\ \mathbf{0}]O^T \\ =& OQ^{-1}O^T - O \begin{bmatrix} Q_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} O^T. \end{aligned}$$

Since Q is symmetric and positive definite, it has a (unique) Cholesky decomposition $Q = LL^T$ where L is lower-triangular with positive diagonal entries. So we can write

$$Q = LL^T = \begin{bmatrix} L_{11} & \mathbf{0} \\ L_{12} & L_{22} \end{bmatrix} \begin{bmatrix} L_{11}^T & L_{12}^T \\ \mathbf{0} & L_{22}^T \end{bmatrix} = \begin{bmatrix} L_{11}L_{11}^T = Q_1 & L_{11}L_{12}^T \\ L_{12}L_{11}^T & * \end{bmatrix}$$

where L_{11}, L_{22} are lower-triangular with positive diagonal entries (and hence invertible), and

$$Q_1^{-1} = L_{11}^{-T} L_{11}^{-1},$$
$$L^{-1} = \begin{bmatrix} L_{11}^{-1} & \mathbf{0} \\ * & L_{22}^{-1} \end{bmatrix}$$

For any m-dimensional vector $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$, where $\boldsymbol{x}_1$ has dimension q and $\boldsymbol{x}_2$ has dimension $m - q$,

$$\begin{aligned} & \boldsymbol{x} \left(Q^{-1} - \begin{bmatrix} Q_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \right) \boldsymbol{x}^T \\ =& \boldsymbol{x} \begin{bmatrix} L_{11}^{-T} & * \\ \mathbf{0} & L_{22}^{-T} \end{bmatrix} \begin{bmatrix} L_{11}^{-1} & \mathbf{0} \\ * & L_{22}^{-1} \end{bmatrix} \boldsymbol{x}^T - \boldsymbol{x} \begin{bmatrix} L_{11}^{-T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} L_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{x}^T \\ =& \|(L_{11}^{-1}\boldsymbol{x}_1, *)\|_2^2 - \|(L_{11}^{-1}\boldsymbol{x}_1, \mathbf{0})\|_2^2 \\ \geq& 0 \end{aligned}$$

So

$$Q^{-1} - \begin{bmatrix} Q_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

is positive semi-definite and so is $A - G(G^T A^{-1} G)^{-1} G^T$. □