Web-based Supplementary Materials for "On latent-variable model misspecification in structural measurement error models for binary response"

Xianzheng Huang* **and Joshua M. Tebbs**

Department of Statistics, University of South Carolina,

Columbia, South Carolina 29208, U.S.A.

**email:* huang@stat.sc.edu

Web Appendix A: Estimation of var$\{\widehat{\boldsymbol{\Omega}}_{(.)}(\lambda) - \widehat{\boldsymbol{\Omega}}_{(.)}(0)\}$ in Section 3

Consider the estimators based on the data with individual responses, $\widehat{\boldsymbol{\Omega}}_I(0)$ and $\widehat{\boldsymbol{\Omega}}_I(\lambda)$. Henceforth, we drop the subscript I for brevity. Recall that $\widehat{\boldsymbol{\Omega}}(0)$ solves (6) and that $\widehat{\boldsymbol{\Omega}}(\lambda)$ solves (7) in the article. Assume that $\boldsymbol{\Omega}(0)$ and $\boldsymbol{\Omega}(\lambda)$ are uniquely defined by

$$E\left[\boldsymbol{\psi}\{Y_i, W_i; \boldsymbol{\Omega}(0), \sigma_U^2\}\right] = \mathbf{0}$$

and

$$E\left[\boldsymbol{\psi}_B\{Y_i, \widetilde{\mathbf{W}}_i(\lambda); \boldsymbol{\Omega}(\lambda), (1+\lambda)\sigma_U^2\}\right] = \mathbf{0},$$

respectively. Approximating the estimators with the use of influence functions associated with (6) and (7), respectively, we write

$$\widehat{\boldsymbol{\Omega}}(0) \approx \boldsymbol{\Omega}(0) + n^{-1}\mathbf{A}_1^{-1}\sum_{i=1}^{n}\boldsymbol{\psi}\{Y_i, W_i; \boldsymbol{\Omega}(0), \sigma_U^2\} \tag{A.1}$$

and

$$\widehat{\boldsymbol{\Omega}}(\lambda) \approx \boldsymbol{\Omega}(\lambda) + n^{-1}\mathbf{A}_2^{-1}\sum_{i=1}^{n}\boldsymbol{\psi}_B\{Y_i, \widetilde{\mathbf{W}}_i(\lambda); \boldsymbol{\Omega}(\lambda), (1+\lambda)\sigma_U^2\}, \tag{A.2}$$

where the matrices

$$\mathbf{A}_1 = E\left[-\frac{\partial}{\partial\boldsymbol{\Omega}^T}\boldsymbol{\psi}\left\{Y_i, W_i; \boldsymbol{\Omega}, \sigma_U^2\right\}\right]$$

and

$$\mathbf{A}_2 = E\left[-\frac{\partial}{\partial\boldsymbol{\Omega}^T}\boldsymbol{\psi}_B\{Y_i, \widetilde{\mathbf{W}}_i(\lambda); \boldsymbol{\Omega}, (1+\lambda)\sigma_U^2\}\right]$$

are evaluated at $\mathbf{\Omega}(0)$ and $\mathbf{\Omega}(\lambda)$, respectively. Subtracting (A.1) from (A.2) gives

$$\begin{aligned}\widehat{\mathbf{\Omega}}(\lambda)-\widehat{\mathbf{\Omega}}(0) &\approx \mathbf{\Omega}(\lambda)-\mathbf{\Omega}(0)+ \\ & \quad n^{-1}\sum_{i=1}^{n}\left[\mathbf{A}_2^{-1}\boldsymbol{\psi}_B\{Y_i,\widetilde{\mathbf{W}}_i(\lambda);\mathbf{\Omega}(\lambda),(1+\lambda)\sigma_U^2\}-\mathbf{A}_1^{-1}\boldsymbol{\psi}\{Y_i,W_i;\mathbf{\Omega}(0),\sigma_U^2\}\right] \\ &= \mathbf{\Omega}(\lambda)-\mathbf{\Omega}(0)+n^{-1}\sum_{i=1}^{n}\mathbf{R}_i, \qquad (A.3)\end{aligned}$$

where $\mathbf{R}_i=\mathbf{A}_2^{-1}\boldsymbol{\psi}_B\{Y_i,\widetilde{\mathbf{W}}_i(\lambda);\mathbf{\Omega}(\lambda),(1+\lambda)\sigma_U^2\}-\mathbf{A}_1^{-1}\boldsymbol{\psi}\{Y_i,W_i;\mathbf{\Omega}(0),\sigma_U^2\}$, for $i=1,2,...,n$. We estimate $\mathbf{R}_i$ by $\widehat{\mathbf{R}}_i=\widehat{\mathbf{A}}_2^{-1}\boldsymbol{\psi}_B\{Y_i,\widetilde{\mathbf{W}}_i(\lambda);\widehat{\mathbf{\Omega}}(\lambda),(1+\lambda)\sigma_U^2\}-\widehat{\mathbf{A}}_1^{-1}\boldsymbol{\psi}\{Y_i,W_i;\widehat{\mathbf{\Omega}}(0),\sigma_U^2\}$, where the matrices

$$\begin{aligned}\widehat{\mathbf{A}}_1 &= n^{-1}\sum_{i=1}^{n}\left[-\frac{\partial}{\partial\mathbf{\Omega}^T}\boldsymbol{\psi}\{Y_i,W_i;\mathbf{\Omega},\sigma_U^2\}\right] \\ \widehat{\mathbf{A}}_2 &= n^{-1}\sum_{i=1}^{n}\left[-\frac{\partial}{\partial\mathbf{\Omega}^T}\boldsymbol{\psi}_B\{Y_i,\widetilde{\mathbf{W}}_i(\lambda);\mathbf{\Omega},(1+\lambda)\sigma_U^2\}\right]\end{aligned}$$

are evaluated at $\widehat{\mathbf{\Omega}}(0)$ and $\widehat{\mathbf{\Omega}}(\lambda)$, respectively. A natural estimator of $\text{var}\{\widehat{\mathbf{\Omega}}(\lambda)-\widehat{\mathbf{\Omega}}(0)\}$, based on the approximation in (A.3), is then given by n^{-1} times the sample covariance matrix of $\widehat{\mathbf{R}}_1,\widehat{\mathbf{R}}_2,...,\widehat{\mathbf{R}}_n$. To construct an estimator for $\text{var}\{\widehat{\mathbf{\Omega}}_R(\lambda)-\widehat{\mathbf{\Omega}}_R(0)\}$, the same argument applies. One simply needs to replace the score functions in (6) and (7) with the score functions in (8) and (9) in the article, respectively, change the index in the summation from i to g, and change the sample size from n to G.

Web Appendix B: Estimation of var$\{\widehat{\mathbf{\Omega}}_R(0)-\widehat{\mathbf{\Omega}}_I(0)\}$ in Section 5

To simplify notation, we relabel the individual-response data after random pooling so that in the first pool, $\mathbf{W}_1^*$ is the collection of the first n_1 W's in the individual-response data, $W_1,W_2,\ldots,W_{n_1}$; and Y_1^* is also based on the first n_1 individual binary responses, i.e., $Y_1^*=\max\{Y_1,Y_2,\ldots,Y_{n_1}\}$. In general, for the gth pool, $\mathbf{W}_g^*=(W_{n_{g-1}+1},W_{n_{g-1}+2},\ldots,W_{n_{g-1}+n_g})$ and $Y_g^*=\max\{Y_{n_{g-1}+1},Y_{n_{g-1}+2},\ldots,Y_{n_{g-1}+n_g}\}$, for $g=2,3,...,G$. For simplicity, we assume that $n_g=k$, for all g, so that $n=kG$.

Using the approximation in (A.1) with grouping within the summation gives

$$\begin{aligned}\widehat{\mathbf{\Omega}}_I(0) &\approx \mathbf{\Omega}_I(0) + n^{-1}\mathbf{A}_1^{-1}\sum_{i=1}^{n}\boldsymbol{\psi}\{Y_i, W_i; \mathbf{\Omega}_I(0), \sigma_U^2\} \\ &= \mathbf{\Omega}_I(0) + G^{-1}\sum_{g=1}^{G} k^{-1}\mathbf{A}_1^{-1}\sum_{j=1}^{k}\boldsymbol{\psi}\{Y_{gj}, W_{gj}; \mathbf{\Omega}_I(0), \sigma_U^2\}. \qquad \text{(B.1)}\end{aligned}$$

Similarly, based on (8), one has

$$\widehat{\mathbf{\Omega}}_R(0) \approx \mathbf{\Omega}_R(0) + G^{-1}\mathbf{A}_3^{-1}\sum_{g=1}^{G}\boldsymbol{\psi}^*\{Y_g^*, \mathbf{W}_g^*; \mathbf{\Omega}_R(0), \sigma_U^2\}, \qquad \text{(B.2)}$$

where the matrix

$$\mathbf{A}_3 = E\left[-\frac{\partial}{\partial \mathbf{\Omega}^T}\boldsymbol{\psi}^*\{Y_g^*, \mathbf{W}_g^*; \mathbf{\Omega}, \sigma_U^2\}\right]$$

is evaluated at $\mathbf{\Omega} = \mathbf{\Omega}_R(0)$. Subtracting (B.1) from (B.2) yields

$$\begin{aligned}\widehat{\mathbf{\Omega}}_R(0) - \widehat{\mathbf{\Omega}}_I(0) &\approx \mathbf{\Omega}_R(0) - \mathbf{\Omega}_I(0) + \\ &G^{-1}\sum_{g=1}^{G}\left[\mathbf{A}_3^{-1}\boldsymbol{\psi}^*\{Y_g^*, \mathbf{W}_g^*; \mathbf{\Omega}_R(0), \sigma_U^2\} - k^{-1}\mathbf{A}_1^{-1}\sum_{j=1}^{k}\boldsymbol{\psi}\{Y_{gj}, W_{gj}; \mathbf{\Omega}_I(0), \sigma_U^2\}\right]. \qquad \text{(B.3)}\end{aligned}$$

Note that the summand in (B.3) denoted by

$$\mathbf{S}_g = \mathbf{A}_3^{-1}\boldsymbol{\psi}^*\{Y_g^*, \mathbf{W}_g^*; \mathbf{\Omega}_R(0), \sigma_U^2\} - k^{-1}\mathbf{A}_1^{-1}\sum_{j=1}^{k}\boldsymbol{\psi}\{Y_{gj}, W_{gj}; \mathbf{\Omega}_I(0), \sigma_U^2\},$$

for $g = 1, 2, ..., G$, are independent. Therefore, a natural estimator for $\text{var}\{\widehat{\mathbf{\Omega}}_R(0) - \widehat{\mathbf{\Omega}}_I(0)\}$, denoted by $\widehat{\mathbf{\Sigma}}$, is given by G^{-1} times the sample covariance matrix of $\widehat{\mathbf{S}}_1, \widehat{\mathbf{S}}_2, ..., \widehat{\mathbf{S}}_G$, where

$$\widehat{\mathbf{S}}_g = \widehat{\mathbf{A}}_3^{-1}\boldsymbol{\psi}^*\{Y_g^*, \mathbf{W}_g^*; \widehat{\mathbf{\Omega}}_R(0), \sigma_U^2\} - k^{-1}\widehat{\mathbf{A}}_1^{-1}\sum_{j=1}^{k}\boldsymbol{\psi}\{Y_{gj}, W_{gj}; \widehat{\mathbf{\Omega}}_I(0), \sigma_U^2\}$$

and the matrix

$$\widehat{\mathbf{A}}_3 = G^{-1}\sum_{g=1}^{G}\left[-\frac{\partial}{\partial \mathbf{\Omega}^T}\boldsymbol{\psi}^*\{Y_g^*, \mathbf{W}_g^*; \mathbf{\Omega}, \sigma_U^2\}\right]$$

is evaluated at $\mathbf{\Omega} = \widehat{\mathbf{\Omega}}_R(0)$.

Under H_{02} : $\mathbf{\Omega}_R(0) = \mathbf{\Omega}_I(0)$, due to the sample-variance construction of $\widehat{\mathbf{\Omega}}$, $\{\widehat{\mathbf{\Omega}}_R(0) - \widehat{\mathbf{\Omega}}_I(0)\}^T\widehat{\mathbf{\Sigma}}^{-1}\{\widehat{\mathbf{\Omega}}_R(0) - \widehat{\mathbf{\Omega}}_I(0)\}$ is a Hotelling's T^2 statistic (Corollary 3.5.1.1 in *Multivariate*

Analysis by Mardia, Kent, and Bibby, 1979). Using the properties of a Hotelling's T^2 statistic, $t_2^* = (G-r)\{r(G-1)\}^{-1}\{\widehat{\boldsymbol{\Omega}}(0) - \widehat{\boldsymbol{\Omega}}\}^T \widehat{\boldsymbol{\Sigma}}^{-1}\{\widehat{\boldsymbol{\Omega}}_R(0) - \widehat{\boldsymbol{\Omega}}_I(0)\}$ follows an $F(r, G-r)$ distribution asymptotically. The quantile-quantile plots (not shown) of the values of t_2^* under H_{02} from our simulation studies in Section 5.2 provide empirical evidence of the claimed asymptotic distribution.