Supplementary Material for “Matched Case-Control Data with a Misclassified Exposure: What can be done with Instrumental Variables?”

CHRISTOPHER M. MANUEL, SAMIRAN SINHA*, SUOJIN WANG

Department of Statistics, Texas A&M University, College Station, Texas 77843-3143

*sinha@stat.tamu.edu

A.1. Identification of the parameters of the model \( \Pr(W = 1|S, X^*, Y = 0, Z) \)

The identification comes from the assumed non-linear structure for \( \Pr(X = 1|S, X^*, Y = 0, Z) \).

Had \( \Pr(X = 1|S, X^*, Y = 0, Z) \) been linear, the parameters would not be identifiable. In short we write \( H(\gamma_0 + \gamma_2^T S + \gamma_3^T X^* + \gamma_4^T Z) \) as \( H(\gamma, S, X^*, Z) \). In our case \( H(\cdot) \) is the logistic function, which is nonlinear.

To see the identifiability issue, we need to show that for every given parameter set \( (\gamma, \alpha_0, \alpha_1) \) if another parameter set \( (\gamma^*, \alpha_0^*, \alpha_1^*) \) satisfies \( \Pr(W = 1|S, X^*, Y = 0, Z; \alpha_0, \alpha_1, \gamma) = \Pr(W = 1|S, X^*, Y = 0, Z; \alpha_0^*, \alpha_1^*, \gamma^*) \) for every choice of \( S, X^* \) and \( Z \), then \( (\gamma^*, \alpha_0^*, \alpha_1^*) = (\gamma, \alpha_0, \alpha_1) \).

To see this, by Equation (3.3) we start with

\[
\alpha_0 + (1 - \alpha_0 - \alpha_1)H(\gamma, S, X^*, Z) = \alpha_0^* + (1 - \alpha_0^* - \alpha_1^*)H(\gamma^*, S, X^*, Z)
\] (A.1)

for every choice of \((S^T, X^{*T}, Z^{*T})^T\). Let \( \gamma^* = -\gamma \), \( \alpha_0^* = 1 - \alpha_1 \) and \( \alpha_1^* = 1 - \alpha_0 \). Then
\(H(\gamma^*, S, X^*, Z) = H(-\gamma, S, X^*, Z) = 1 - H(\gamma, S, X^*, Z)\) and
\[
\alpha_0^* + (1 - \alpha_0^* - \alpha_1^*)H(\gamma^*, S, X^*, Z) = (1 - \alpha_1) + (1 - 1 + \alpha_1 - 1 + \alpha_0)H(-\gamma, S, X^*, Z)
\]
\[
= (1 - \alpha_1) + (-1 + \alpha_0 + \alpha_1)\{1 - H(\gamma, S, X^*, Z)\}
\]
\[
= (1 - \alpha_1) + (-1 + \alpha_0 + \alpha_1) - (1 + \alpha_0 + \alpha_1)H(\gamma, S, X^*, Z)
\]
\[
= \alpha_0 + (1 - \alpha_0 - \alpha_1)H(\gamma, S, X^*, Z).
\]

On the other hand, under the monotonicity restriction \(\alpha_0 + \alpha_1 < 1\), if \(\alpha_1^* = 1 - \alpha_0\) and \(\alpha_0^* = 1 - \alpha_1\), then \(\alpha_0^* + \alpha_1^* = (1 - \alpha_1 + 1 - \alpha_0) = 1 + (1 - \alpha_0 - \alpha_1) > 1\). Hence, this particular choice of \(\alpha_0^*, \alpha_1^*\) does not satisfy the restriction, and is not a cause of concern anymore.

Finally, we need to check if there is any other choice of \((\alpha_0^*, \alpha_1^*, \gamma^*)\) that satisfies (A.1). Suppose that there exists \((\alpha_0^*, \alpha_1^*, \gamma^*)\) that satisfies (A.1) for every choice of \(S, X^*\) and \(Z\). This implies that for every \((S_k, X_k^*, Z_k)\), \(k = 1, 2, \ldots\),
\[
\alpha_0^* + (1 - \alpha_0^* - \alpha_1^*)H(\gamma^*, S_k, X_k^*, Z_k) = \alpha_0 + (1 - \alpha_0 - \alpha_1)H(\gamma, S_k, X_k^*, Z_k).
\]

Since \(1 - \alpha_0^* - \alpha_1^* > 0\) and \(1 - \alpha_0 - \alpha_1 > 0\), it is readily seen that each element of \((\gamma_1^*, \gamma_2^*)\) must have the same sign as the corresponding element of \((\gamma_1, \gamma_2)\). By letting \(T = \gamma_0 + \gamma_1^T S + \gamma_2^T X^* + \gamma_3^T Z \to -\infty\) (and then \(T^* = \gamma_0^* + \gamma_1^T S + \gamma_2^T X^* + \gamma_3^T Z \to -\infty\) also), it is clear that \(\alpha_0^* = \alpha_0\). Likewise, due to the nonlinearity of \(H(\cdot)\), \(\alpha_1^* = \alpha_1\). This leads to \(T^* = T\) and thus \(\gamma^* = \gamma\), showing the identifiability of these parameters.

A.2. Proof of Lemma 1

Because of the logistic model assumption and the assumption on \(W\) and \(X^*\) we can write
\[
1 - \Pr(Y = 0|S, W, X, X^*, Z) = \Pr(Y = 1|S, W, X, X^*, Z)
\]
\[
= \Pr(Y = 1|S, X, Z)
\]
\[
= \exp\{g_0(S) + \beta_1 X + \beta_2^T Z\}\Pr(Y = 0|S, X, Z),
\]
where $g_0(\cdot)$ is given in Model (2.1). We now consider

\[
\begin{align*}
\text{pr}(Y = 1|S, W, X^*, Z) &= \sum_{x=0,1} \text{pr}(Y = 1|S, W, X = x, X^*, Z) \text{pr}(X = x|S, W, X^*, Z) \\
&= \sum_{x=0,1} \text{pr}(Y = 1|S, X = x, Z) \text{pr}(X = x|S, W, X^*, Z) \\
&= \sum_{x=0,1} \exp\{g_0(S_i) + \beta_1 x + \beta_2^T Z\} \text{pr}(Y = 0|S, X = x, Z) \text{pr}(X = x|S, W, X^*, Z) \\
&= \sum_{x=0,1} \exp\{g_0(S_i) + \beta_1 x + \beta_2^T Z\} \text{pr}(X = x|S, W, X^*, Y = 0, Z) \text{pr}(Y = 0|S, W, X^*, Z) \\
&= \text{pr}(Y = 0|S, W, X^*, Z) \sum_{x=0,1} \exp\{g_0(S) + \beta_1 x + \beta_2^T Z\} \text{pr}(X = x|S, W, X^*, Y = 0, Z) \\
&= \text{pr}(Y = 0|S, W, X^*, Z) \exp\{g_0(S) + \beta_2^T Z\} \{\exp(\beta_1) \text{pr}(X = 1|S, W, X^*, Y = 0, Z) \\
& \quad + \text{pr}(X = 0|S, W, X^*, Y = 0, Z)\} \\
& \equiv \text{pr}(Y = 0|S, W, X^*, Z) \exp\{g_0(S) + \beta_2^T Z + g_1(\beta_1, S_i, W, X^*, Z, \gamma, \eta)\},
\end{align*}
\]

where the expression of $g_1(\beta_1, S, W, X^*, Z, \gamma, \eta)$ is obtained after plugging the expression for

\[
\text{pr}(X = 1|S, W, X^*, Y = 0, Z) \text{ and } \text{pr}(X = 0|S, W, X^*, Y = 0, Z) \text{ from Equations (3.4) and (3.5). In particular,}
\]

\[
\begin{align*}
\exp\{g_1(\beta_1, S, W = 1, X^*, Z, \gamma, \eta)\} &= \exp(\beta_1) \text{pr}(X = 1|S, W = 1, X^*, Y = 0, Z) + \text{pr}(X = 0|S, W = 1, X^*, Y = 0, Z) \\
&= \exp(\beta_1) \frac{(1 - \alpha_1) H(\gamma, S, X^*, Z)}{\alpha_0 + (1 - \alpha_0 - \alpha_1) H(\gamma, S, X^*, Z)} + 1 - \frac{(1 - \alpha_1) H(\gamma, S, X^*, Z)}{\alpha_0 + (1 - \alpha_0 - \alpha_1) H(\gamma, S, X^*, Z)} \\
&= \frac{\exp(\beta_1)(1 - \alpha_1) H(\gamma, S, X^*, Z) + \alpha_0 \{1 - H(\gamma, S, X^*, Z)\}}{\alpha_0 + (1 - \alpha_0 - \alpha_1) H(\gamma, S, X^*, Z)}, \quad (A.2)
\end{align*}
\]

\[
\begin{align*}
\exp\{g_1(\beta_1, S, W = 0, X^*, \gamma, \eta)\} &= \exp(\beta_1) \text{pr}(X = 1|S, W = 0, X^*, Y = 0, Z) + \text{pr}(X = 0|S, W = 0, X^*, Y = 0, Z) \\
&= \exp(\beta_1) \frac{\alpha_1 H(\gamma, S, X^*, Z)}{1 - \alpha_0 - (1 - \alpha_0 - \alpha_1) H(\gamma, S, X^*, Z)} + 1 - \frac{\alpha_1 H(\gamma, S, X^*, Z)}{1 - \alpha_0 - (1 - \alpha_0 - \alpha_1) H(\gamma, S, X^*, Z)} \\
&= \frac{\exp(\beta_1) \alpha_1 H(\gamma, S, X^*, Z) + (1 - \alpha_0) \{1 - H(\gamma, S, X^*, Z)\}}{1 - \alpha_0 - (1 - \alpha_0 - \alpha_1) H(\gamma, S, X^*, Z)}. \quad (A.3)
\end{align*}
\]
A.3. Proof of Theorem 1

Collecting $S_\gamma(\gamma, \eta), S_\eta(\gamma, \eta), S_{\beta_1}(\beta, \gamma, \eta), S_{\beta_2}(\beta, \gamma, \eta)$ together and letting $\theta = (\gamma^T, \eta^T, \beta_1, \beta_2^T)^T$ and $\hat{\theta} = (\hat{\gamma}^T, \hat{\eta}^T, \hat{\beta}_1, \hat{\beta}_2^T)^T$, we can write

$$\sqrt{n}(\hat{\theta} - \theta) = A^{-1} \sum_{i=1}^{n} U_i + o_p(1),$$

where $U_i$'s are iid and mean zero and finite variance random vectors. $A = -E(\partial U_i / \partial \theta)$. By the Central Limit Theorem we obtain the asymptotic normality of $\hat{\theta}$, and the asymptotic variance of $\sqrt{n} \hat{\theta}$ is $A^{-1} \text{var}(U_1) A^{-T}$. This asymptotic variance can be consistently estimated by $A^{-1} (\sum_{i=1}^{n} \hat{U}_i \hat{U}_i^T / n) A^{-T}$ with $A = -(1/n) \sum_{i=1}^{n} \partial \hat{U}_i / \partial \theta$ and $\hat{U}_i$ being $U_i$ with $\theta$ replaced by $\hat{\theta}$.

A.4. Proof of Lemma 2

Part i) of Lemma 2

$$\text{pr}(Y = 1 | S, X^*, Z) = \sum_x \text{pr}(Y = 1 | S, X = x, X^*, Z) \text{pr}(X = x | S, X^*, Z)$$

$$= \sum_x \exp\{g_0(S) + \beta_1 x + \beta_2^T Z\} \text{pr}(Y = 0 | S, X = x, X^*, Z) \text{pr}(X = x | S, X^*, Z)$$

$$= \sum_x \exp\{g_0(S) + \beta_1 x + \beta_2^T Z\} \text{pr}(X = x | S, X^*, Y = 0, Z) \text{pr}(Y = 0 | S, X^*, Z)$$

$$= \text{pr}(Y = 0 | S, X^*, Z) \left\{ \exp\{g_0(S) + \beta_2^T Z\} \{1 - H(\gamma, S, X^*, Z)\} \right\}$$

$$+ \exp\{g_0(S) + \beta_1 + \beta_2^T Z\} H(\gamma, S, X^*, Z)\}$$

$$= \text{pr}(Y = 0 | S, X^*, Z) \exp\{g_0(S) + \beta_2^T Z\}$$

$$\times \{1 - H(\gamma, S, X^*, Z) + \exp(\beta_1) H(\gamma, S, X^*, Z)\}.$$
where
\[
g_2(\gamma, \beta_1, S, X^*, Z) = \log \{1 - H(\gamma, S, X^*, Z) + \exp(\beta_1)H(\gamma, S, X^*, Z)\}.
\]

**Part ii) of Lemma 2**

\[
\Pr(X = 1|S, X^*, Z, Y = 1) = \frac{\Pr(Y = 1|S, X = 1, X^*, Z)\Pr(X = 1|S, X^*, Z)}{\Pr(Y = 1|S, X^*, Z)}
\]

\[
= \frac{\exp\{g_0(S) + \beta_1 + \beta_2^T Z\} \Pr(Y = 0|S, X = 1, X^*, Z)\Pr(X = 1|S, X^*, Z)}{\Pr(Y = 1|S, X^*, Z)}
\]

\[
= \frac{\exp\{g_0(S) + \beta_1 + \beta_2^T Z\} \Pr(X = 1|S, X^*, Y = 0, Z)\Pr(Y = 0|S, X^*, Z)}{\Pr(Y = 1|S, X^*, Z)}
\]

\[
= \frac{\exp\{g_0(S) + \beta_1 + \beta_2^T Z\}H(\gamma, S, X^*, Z)}{\Pr(Y = 1|S, X^*, Z)}
\]

\[
= \frac{\exp\{g_0(S) + \beta_1 + \beta_2^T Z\}H(\gamma, S, X^*, Z)}{\Pr(Y = 1|S, X^*, Z)}
\]

\[
= \frac{\exp\{g_0(S) + \beta_1 + \beta_2^T Z\}H(\gamma, S, X^*, Z)}{1 - H(\gamma, S, X^*, Z) + \exp(\beta_1)H(\gamma, S, X^*, Z)}
\]

\[
= \frac{\exp(\gamma_0 + \beta_1 + \gamma_1^T S + \gamma_2^T X^* + \gamma_3^T Z)}{1 + \exp(\gamma_0 + \beta_1 + \gamma_1^T S + \gamma_2^T X^* + \gamma_3^T Z)}
\]

\[
= H(\gamma_0 + \beta_1 + \gamma_1^T S + \gamma_2^T X^* + \gamma_3^T Z).
\]

**Part iii) of Lemma 2**

\[
\Pr(W = 1|S, X^*, Y = 1, Z)
\]

\[
= \Pr(W = 1|S, X = 0, X^*, Y = 1, Z)\Pr(X = 0|S, X^*, Y = 1, Z)
\]

\[
+ \Pr(W = 1|S, X = 1, X^*, Y = 1, Z)\Pr(X = 1|S, X^*, Y = 1, Z)
\]

\[
= \Pr(W = 1|X = 0)\Pr(X = 0|S, X^*, Y = 1, Z)
\]

\[
+ \Pr(W = 1|X = 1)\Pr(X = 1|S, X^*, Y = 1, Z)
\]

\[
= \alpha_0 \{1 - \Pr(X = 1|S, X^*, Y = 1, Z)\} + (1 - \alpha_1)\Pr(X = 1|S, X^*, Y = 1, Z)
\]

\[
= \alpha_0 + (1 - \alpha_0 - \alpha_1)\Pr(X = 1|S, X^*, Y = 1, Z)
\]

\[
= \alpha_0 + (1 - \alpha_0 - \alpha_1)H(\gamma_0 + \beta_1 + \gamma_1^T S + \gamma_2^T X^* + \gamma_3^T Z).
\]