Supplemental Appendix
Semiparametric inference under a discrete choice model for nonmonotone missing not at random data
Eric J. Tchetgen Tchetgen, Linbo Wang, BaoLuo Sun
Department of Biostatistics, Harvard University

1 Sensitivity analysis for CCMV

Identification conditions such as CCMV are not generally empirically testable and therefore, it is important that inferences in a given analysis are assessed for sensitivity to violation of such assumptions. Specifically, a violation of the CCMV assumption can occur if for some \( r \),

\[ R \not \perp L(-r) | L(r), R \in \{1, r\}, \]

which can be encoded by specifying the degree of departure from the identifying assumption, on the odds ratio scale using the selection bias function:

\[
\theta_r (L(-r), L(r)) = \frac{\pi_r (L(r), L(-r)) \pi_1 (L(r), L(-r) = 0)}{\pi_1 (L(r), L(-r)) \pi_r (L(r), L(-r) = 0)}.
\]

CCMV corresponds to the null \( \theta_r (L(-r), L(r)) = 1 \) for all \( r \), and \( \theta_r (L(-r), L(r)) \neq 1 \) for some \( r \) indicates violation of the assumption. The function \( \theta_r (\cdot, \cdot) \) is not nonparametrically identified from the observed data. Therefore we propose that one may specify a functional form for \( \theta_r (\cdot, \cdot) \) for use in a sensitivity analysis in the spirit of Robins et al (1999). Hereafter, suppose that one has specified functions \( \theta = \{\theta_r : r\} \). For such specification, we describe IPW, PM and DR estimation incorporating a non-null \( \theta_r \).

For IPW estimation, we propose to modify \( W_r \) of Section 5 as follows. Let \( W_r (G_r; \alpha, \theta_r) = G_r \times [1 \{R = r\} - 1 \{R = 1\} \theta_r (L) \Pi_r (\alpha) / \Pi_1 (\alpha)] \), and denote by \( \hat{\alpha} (\theta) \) the solution to \( \mathbb{P}_n W_r (G_r; \hat{\alpha} (\theta), \theta_r) = 0 \), then a consistent IPW estimator \( \hat{\beta}_{ipw} (\theta) \) solves equation (10) in the main text with \( \Pi_1 \) replaced
by $\Pi_1^\ast (\hat{\alpha} (\theta)) = \left\{ 1 + \sum_{r \neq 1} \theta_r (L) \Pi_r (\hat{\alpha} (\theta)) / \Pi_1 (\hat{\alpha} (\theta)) \right\}^{-1}$.

Likewise, PM estimation hinges on the following expression

$$E \{ U(L; \beta) | R = r, L(r); \tilde{\eta}, \theta \}$$

$$= \frac{\int \theta_r (L(-r), L(r)) U(l(-r), L(r)) ; \beta f (l(-r), L(r)) | R = 1; \eta) d\mu (l(-r))}{\int \theta_r (L(-r), L(r)) f (l(-r), L(r)) | R = 1; \eta) d\mu (l(-r))}$$

which may be used in place of $E \{ U(L; \beta) | R = 1, L(r); \tilde{\eta} \}$ in equation (12), which in turn may be used to obtain the PM estimator $\hat{\beta}_{pm} (\theta)$. Finally, for a given value of $\theta$, the DR estimator $\hat{\beta}_{dr} (\theta)$ solves equation (14) with $V (\hat{\beta}_{dr}, \tilde{\alpha}, \tilde{\eta})$ replaced by

$$V (\beta, \hat{\alpha} (\theta), \tilde{\eta}; \theta) = \left\{ \frac{1 (R = 1)}{\Pi_1^\ast (\hat{\alpha} (\theta))} U(L; \beta) \right\}$$

$$- \frac{1 (R = 1)}{\Pi_1^\ast (\hat{\alpha} (\theta))} \sum_{r \pm 1} \Pi_r^\ast (\hat{\alpha} (\theta)) E [U(L; \beta) | L(r), R = r; \tilde{\eta}, \theta]$$

$$+ \sum_{r \pm 1} I (R = r) E [U(L; \beta) | L(r), R = r; \tilde{\eta}, \theta],$$

where

$$\Pi_r^\ast (\hat{\alpha} (\theta)) = \frac{\theta_r (L) \Pi_r (\hat{\alpha} (\theta)) / \Pi_1 (\hat{\alpha} (\theta))}{\left\{ 1 + \sum_{r' \neq 1} \theta_{r'} (L) \Pi_{r'} (\hat{\alpha} (\theta)) / \Pi_1 (\hat{\alpha} (\theta)) \right\}}.$$

A sensitivity analysis then entails reporting $\hat{\beta}_{ipw} (\theta)$, $\hat{\beta}_{pm} (\theta)$ or $\hat{\beta}_{dr} (\theta)$ for a range of values of $\theta$.

### 2 Proof of Lemmas

**Proof of Lemma 1:** The result follows from the following generalized odds ratio representation of the joint likelihood of $f(R, L)$ (see Chen, 2007 and Tchetgen Tchetgen et al, 2010)

$$f(R, L) = \frac{\int f (R|L = 0) f(L|R = 1) OR (R, L)}{\int f (r^*|L = 0) f(l^*|R = 1) OR (r^*, l^*) d\mu (r^*, l^*)},$$
provided that \( \int \int f(r^*|L = 0) f(l^*|R = 1) \text{OR}(r^*, l^*) d\mu(r^*, l^*) < \infty \), where the generalized odds ratio function \( \text{OR}(R, L) \) is defined as

\[
\text{OR}(R, L) = \frac{f(R, L) f(R = 1, L = 0)}{f(R = 1, L) f(R, L = 0)}.
\]

Then

\[
\frac{\int \int f(r^*|L = 0) f(l^*|R = 1) \text{OR}(r^*, l^*) d\mu(r^*, l^*)}{\int \int f(r^*|L = 0) f(l^*|R = 1) \text{OR}(r^*, l^*) d\mu(r^*, l^*)} = \frac{\int \int \frac{f(R|L = 0)}{f(R = 1|L = 0)} \text{OR}(R, L) f(L|R = 1) d\mu(r^*, l^*)}{\int \int \prod_{r \neq 1} \text{Odds}_r(L_I(L = r)) f(L|R = 1) d\mu(r^*, l^*)}
\]

proving the result.

**Proof of Lemma 2:** The complete-case joint distribution \( f(L|R = 1) \) is nonparametrically just-identified under assumption (1). Furthermore, pairwise MAR implies that \( \text{Odds}_r(L) = \text{Odds}_r(L(r)) \) is nonparametrically just-identified from data \( \{(R, L(r)) : R \in \{1, r\}\} \), because \( L_{(r)} \) is MAR conditional on \( L(r) \) and \( R \in \{1, r\} \). Specifically,

\[
\text{Pr}\{R = r|L, R \in \{1, r\}\} = \frac{\text{Pr}\{R = r, L\}}{\text{Pr}\{L, R \in \{1, r\}\}} = \frac{\text{Odds}_r(L_{(r)}) f(L|R = 1) f(L|R = 1)}{\text{Odds}_r(L_{(r)}) f(L|R = 1) f(L|R = 1) + f(L|R = 1) f(L|R = 1)} = \frac{\text{Odds}_r(L_{(r)})}{\text{Odds}_r(L_{(r)}) + 1},
\]

proving the result.

**Proof of Theorem 3:** The result essentially follows from the following DR property of \( V(\beta, \alpha, \eta) \).

Let \( V(\beta, \alpha^*, \eta_0) \) denote the estimating function evaluated at the incorrect \( \Pi_r \) and true \( E[U(L; \beta)|L_{(r)}, R =
\[ E \{ V(\beta, \alpha, \eta^*) \} = E \{ V(\beta_0, \alpha_0, \eta^*) \} = 0. \]

First, note that under \( \mathcal{M}_R \), \( \tilde{\alpha} \to \alpha_0 \) and \( \tilde{\eta} \to \eta^* \) in probability, then
\[
\begin{align*}
\Pr_n V(\beta_0, \tilde{\alpha}, \tilde{\eta}) &\to E \{ V(\beta_0, \alpha_0, \eta^*) \} \text{ in probability by Continuous Mapping Theorem and the Law of Large Numbers. We also have that}
\end{align*}
\]

By the same token, under \( \mathcal{M}_L \), \( \tilde{\alpha} \to \alpha^* \) and \( \tilde{\eta} \to \eta_0 \) in probability, then
\[
\begin{align*}
\Pr_n V(\beta_0, \tilde{\alpha}, \tilde{\eta}) &\to E \{ V(\beta_0, \alpha^*, \eta_0) \} . \text{ Next we show that } E \{ V(\beta_0, \alpha^*, \eta_0) \} = 0. \text{ Note that for all } \alpha
\end{align*}
\]

\[
\frac{1}{\Pi_1(\alpha)} = 1 + \sum_{r \neq 1} \frac{\Pi_r(\alpha)}{\Pi_1(\alpha)} = 1 + \sum_{r \neq 1} \text{Odds}_r(L(r); \alpha) .
\]
Then we have that

\[
E \left( V(\beta, \alpha_0, \eta^*) \right) = E \left\{ \frac{1}{\Pi_1(\alpha^*)} \left\{ U(L; \beta_0) - \sum_{r \neq 1} \Pi_r(\alpha^*) E \left[ U(L; \beta)|L(r), R = 1; \eta_0 \right] \right\} \right.
\]

\[+ \sum_{r \neq 1} 1(R = r) E \left[ U(L; \beta)|L(r), R = 1; \eta_0 \right] \]

\[= E \left\{ 1(R = 1) \left\{ \frac{U(L; \beta_0)}{\Pi_1(\alpha^*)} - \sum_{r \neq 1} \frac{\Pi_r(\alpha^*)}{\Pi_1(\alpha^*)} E \left[ U(L; \beta)|L(r), R = 1; \eta_0 \right] \right\} \right.
\]

\[+ \sum_{r \neq 1} 1(R = r) E \left[ U(L; \beta)|L(r), R = 1; \eta_0 \right] \]

\[= E \left\{ \sum_{r \neq 1} \text{Odds}_r (L(r); \alpha^*) \left( E \left[ U(L; \beta_0)|R = 1, L(r) \right] - E \left[ U(L; \beta)|L(r), R = 1; \eta_0 \right] \right) \right\}
\]

\[= E \left\{ \sum_{r \neq 1} \frac{1}{\Pi_1(\alpha^*)} \left( U(L; \beta_0) - \sum_{r \neq 1} \frac{1}{\Pi_r(\alpha^*)} E \left[ U(L; \beta)|L(r), R = 1; \eta_0 \right] \right) \right\}
\]

\[= E \left\{ 1(R = 1) U(L; \beta_0) + \sum_{r \neq 1} 1(R = r) E \left[ U(L; \beta)|L(r), R = 1; \eta_0 \right] \right\}
\]

\[= E \left\{ 1(R = 1) U(L; \beta_0) + \sum_{r \neq 1} 1(R = r) E \left[ U(L; \beta)|L(r), R = 1; \eta_0 \right] \right\}
\]

\[= E \left\{ 1(R = 1) U(L; \beta_0) + \sum_{r \neq 1} 1(R = r) E \left[ U(L; \beta)|L(r), R = r \right] \right\}
\]

\[= E \left\{ 1(R = 1) E[U(L; \beta_0)|R = 1] + \sum_{r \neq 1} 1(R = r) E[U(L; \beta)|R = r] \right\}
\]

\[= E[U(L; \beta_0)] = 0
\]

proving the result.
Proof of Corollary 4: \( E(V(\beta, \alpha, \eta)) \) can be written

\[
E(V(\beta, \alpha, \eta)) = \sum_{r \neq 1} \frac{1}{\Pi_1(\alpha)} E(U(L; \beta_0) - \sum_{r \neq 1} \frac{1}{\Pi_1(\alpha)} E[U(L; \beta)|L(r), R = 1; \eta]
\]

\[
+ \sum_{r \neq 1} 1(R = r) E[U(L; \beta)|L(r), R = 1; \eta] + 1(R = 1) U(L; \beta_0)
\]

\[
= \sum_{r \neq 1} \frac{1}{\Pi_1(\alpha)} E(U(L; \beta_0) - \sum_{r \neq 1} \frac{1}{\Pi_1(\alpha)} E[U(L; \beta)|L(r), R = 1; \eta]
\]

\[
+ \sum_{r \neq 1} 1(R = r) \left\{ E[U(L; \beta)|L(r), R = 1; \eta] - U(L; \beta_0) \right\} + U(L; \beta_0)
\]

\[
= \sum_{r \neq 1} \{ 1(R = 1) \text{Odds}_r(L(r); \alpha) - 1(R = r) \} \left\{ U(L; \beta_0) - E[U(L; \beta)|L(r), R = 1; \eta] \right\}
\]

Under \( \mathcal{M}_R(r) \), we have that \( \text{Odds}_r(L(r); \tilde{\alpha}) \to \text{Odds}_r(L(r); \alpha_0) \) in probability, and

\[
E \left\{ \left\{ 1(R = 1) \text{Odds}_r(L(r); \alpha_0) - 1(R = r) \right\} \left\{ U(L; \beta_0) - E[U(L; \beta)|L(r), R = 1; \eta^*] \right\} \right\}
\]

\[
= E \left\{ \left\{ 1(R = 1) \frac{\Pi_r}{\Pi_1} - 1(R = r) \right\} \left\{ U(L; \beta_0) - E[U(L; \beta)|L(r), R = 1; \eta^*] \right\} \right\}
\]

\[
= E \left\{ \left\{ \Pi_r - E[1(R = r) \mid L] \right\} \left\{ U(L; \beta_0) - E[U(L; \beta)|L(r), R = 1; \eta^*] \right\} \right\}
\]

\[
= 0
\]

Likewise, under \( \mathcal{M}_L(r) \), we have that \( E[U(L; \beta)|L(r), R = 1; \tilde{\eta}] \to E[U(L; \beta)|L(r), R = 1; \eta_0] \) in probability, and
\begin{align*}
E \left\{ 1 \left( R = 1 \right) \text{Odds}_r \left( L(r); \alpha^* \right) - 1 \left( R = r \right) \right\} \left\{ U(L; \beta_0) - E \left[ U(L; \beta) | L(r), \ R = 1; \eta_0 \right] \right\} \\
= E \left\{ 1 \left( R = 1 \right) \text{Odds}_r \left( L(r); \alpha^* \right) \right\} \left\{ E \left\{ U(L; \beta_0) | R = 1, L(r) \right\} - E \left[ U(L; \beta) | L(r), R = 1; \eta_0 \right] \right\} \\
- E \left\{ \left\{ 1 \left( R = r \right) \right\} \left\{ E \left\{ U(L; \beta_0) | R = r, L(r) \right\} - E \left[ U(L; \beta) | L(r), R = 1; \eta_0 \right] \right\} \right\} \\
= -E \left\{ \left\{ 1 \left( R = r \right) \right\} \left\{ E \left\{ U(L; \beta_0) | R = 1, L(r) \right\} - E \left[ U(L; \beta) | L(r), R = 1; \eta_0 \right] \right\} \right\} \\
= 0
\end{align*}

proving the result.
Table S1: Monte Carlo results of the IPW, PM and DR estimators: bias, standard error and root mean squared error. The true value of $\beta$ is 0.634, and the sample size is 2000.

<table>
<thead>
<tr>
<th></th>
<th>bth*</th>
<th>nrm</th>
<th>ccm</th>
<th>bad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias(SE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPW</td>
<td>-0.004(0.002)</td>
<td>-0.004(0.002)</td>
<td>-0.641(0.012)</td>
<td>-0.641(0.012)</td>
</tr>
<tr>
<td>PM</td>
<td>-0.002(0.001)</td>
<td>-0.367(0.002)</td>
<td>-0.002(0.001)</td>
<td>-0.367(0.002)</td>
</tr>
<tr>
<td>DR</td>
<td>-0.002(0.002)</td>
<td>-0.006(0.002)</td>
<td>-0.002(0.002)</td>
<td>-0.371(0.003)</td>
</tr>
<tr>
<td>RMSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPW</td>
<td>0.072</td>
<td>0.072</td>
<td>0.748</td>
<td>0.748</td>
</tr>
<tr>
<td>PM</td>
<td>0.046</td>
<td>0.373</td>
<td>0.046</td>
<td>0.373</td>
</tr>
<tr>
<td>DR</td>
<td>0.048</td>
<td>0.057</td>
<td>0.057</td>
<td>0.385</td>
</tr>
</tbody>
</table>

*: bth: both models correct; nrm: nonresponse model correct; ccm: complete-case model correct; bad: both models incorrect.

3  Additional Simulation Results

Table S1 shows Monte Carlo results comparing the proposed large sample estimator of standard deviation (and corresponding coverage probabilities of Wald 95% confidence intervals) of IPW, PM and DR estimators of $\beta$ to corresponding Monte Carlo standard deviations.
References

