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,

$$R \not\perp\!\!\!\perp L_{(-r)}|L_{(r)}, R \in \{1, r\},$$

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

$$\theta_r\left(L_{(-r)}, L_{(r)}\right) = \frac{\pi_r\left(L_{(r)}, L_{(-r)}\right)\pi_1\left(L_{(r)}, L_{(-r)} = 0\right)}{\pi_1\left(L_{(r)}, L_{(-r)}\right)\pi_r\left(L_{(r)}, L_{(-r)} = 0\right)}.$$

CCMV corresponds to the null $\theta_r\left(L_{(-r)}, L_{(r)}\right) = 1$ for all r, and $\theta_r\left(L_{(-r)}, L_{(r)}\right) \neq 1$ 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 θ_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 $\widehat{\alpha}(\theta)$ the solution to $\mathbb{P}_n W_r(G_r; \widehat{\alpha}(\theta), \theta_r) = 0$, then a consistent IPW estimator $\widehat{\beta}_{ipw}(\theta)$ solves equation (10) in the main text with Π_1 replaced by $\Pi_1^*(\widehat{\alpha}(\theta)) = \left\{1 + \sum_{r\neq 1} \theta_r(L)\,\Pi_r(\widehat{\alpha}(\theta))/\Pi_1(\widehat{\alpha}(\theta))\right\}^{-1}$

Likewise, PM estimation hinges on the following expression

$$
\begin{aligned}
&E\left\{U(L;\beta)|R=r, L_{(r)}; \widetilde{\eta}, \theta\right\} \\
&= \frac{\int \theta_r\left(L_{(-r)}, L_{(r)}\right) U(l_{(-r)}, L_{(r)}); \beta) f\left(l_{(-r)}, L_{(r)})|R=1;\eta\right) d\mu\left(l_{(-r)}\right)}{\int \theta_r\left(L_{(-r)}, L_{(r)}\right) f\left(l_{(-r)}, L_{(r)})|R=1;\eta\right) d\mu\left(l_{(-r)}\right)}
\end{aligned}
$$

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

$$
\begin{aligned}
V(\beta, \widehat{\alpha}(\theta), \widetilde{\eta}; \theta) = &\left\{\frac{1(R=1)}{\Pi_1^*(\widehat{\alpha}(\theta))} U(L;\beta)\right\} \\
&- \frac{1(R=1)}{\Pi_1^*(\widehat{\alpha}(\theta))} \sum_{r\pm 1} \Pi_r^*(\widehat{\alpha}(\theta))\, E\left[U(L;\beta)|L_{(r)}, R=r; \widetilde{\eta}, \theta\right] \\
&+ \sum_{r\pm 1} I(R=r)\, E\left[U(L;\beta)|L_{(r)}, R=r; \widetilde{\eta}, \theta\right],
\end{aligned}
$$

where

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

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 θ.

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)

$$
f(R, L) = \frac{f(R|L=0)\, f(L|R=1)\mathrm{OR}(R, L)}{\iint f(r^*|L=0)\, f(l^*|R=1)\mathrm{OR}(r^*, l^*)\, d\mu(r^*, l^*)},
$$

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

$$\mathrm{OR}(R,L) = \frac{f(R,L)\, f(R=1,L=0)}{f(R=1,L)\, f(R,L=0)}.$$

Then

$$\begin{aligned}
&\frac{f(R|L=0)\, f(L|R=1)\mathrm{OR}(R,L)}{\iint f(r^*|L=0)\, f(l^*|R=1)\mathrm{OR}(r^*,l^*)\, d\mu(r^*,l^*)} \\
&= \frac{\frac{f(R|L=0)}{f(R=1|L=0)}\mathrm{OR}(R,L)\, f(L|R=1)}{\iint \frac{f(r^*|L=0)}{f(R=1|L=0)}\mathrm{OR}(r^*,l^*)\, f(l^*|R=1) d\mu(r^*,l^*)} \\
&= \frac{\prod_{r\neq 1}\mathrm{Odds}_r(L)^{I(R=r)}\, f(L|R=1)\, f(L|R=1)}{\iint \prod_{r\neq 1}\mathrm{Odds}_r(l^*)^{I(r^*=r)}\, f(l^*|R=1)\, d\mu(r^*,l^*)}
\end{aligned}$$

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 $\mathrm{Odds}_r(L) = \mathrm{Odds}_r\left(L_{(r)}\right)$ is nonparametrically just-identified from data $\left\{(R, L_{(R)}) : R \in \{1, r\}\right\}$, because $L_{(-r)}$ is MAR conditional on $L_{(R)}$ and $R \in \{1, r\}$. Specifically,

$$\begin{aligned}
&\Pr\{R=r|L, R\in\{1,r\}\} \\
&= \frac{\Pr\{R=r, L\}}{\Pr\{L, R\in\{1,r\}\}} \\
&= \frac{\mathrm{Odds}_r\left(L_{(r)}\right) f(L|R=1)\, f(L|R=1)}{\mathrm{Odds}_r\left(L_{(r)}\right) f(L|R=1)\, f(L|R=1) + f(L|R=1)\, f(L|R=1)} \\
&= \frac{\mathrm{Odds}_r\left(L_{(r)}\right)}{\mathrm{Odds}_r\left(L_{(r)}\right)+1},
\end{aligned}$$

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 Π_r and true $E[U(L;\beta)|L_{(r)}, R =$

1] for all r. Likewise let $V(\beta, \alpha_0, \eta^*)$ for the opposite setting. DR property holds if $E\{V(\beta_0, \alpha^*, \eta_0)\} = E\{V(\beta_0, \alpha_0, \eta^*)\} = 0$. First, note that under $\mathcal{M}_R$, $\widetilde{\alpha} \to \alpha_0$ and $\widetilde{\eta} \to \eta^*$ in probability, then $\mathbb{P}_n V(\beta_0, \widetilde{\alpha}, \widetilde{\eta}) \to E\{V(\beta_0, \alpha_0, \eta^*)\}$ in probability by Continuous Mapping Theorem and the Law of Large Numbers. We also have that

$$
\begin{aligned}
E(V(\beta, \alpha_0, \eta^*)) &= E\left\{\frac{1(R=1)}{\Pi_1(\alpha_0)} U(L;\beta_0)\right. \\
&\left. - \sum_{r \neq 1} \left(\frac{1(R=1)\Pi_r(\alpha_0)}{\Pi_1(\alpha)} - 1(R=r)\right) E\left[U(L;\beta)|L_{(r)}, R=1;\eta^*\right]\right\} \\
&= E\left\{\frac{E\{1(R=1)|L\}}{\Pi_1(\alpha_0)} U(L;\beta_0)\right. \\
&- \sum_{r \neq 1} \underbrace{\left(\frac{E\{1(R=1)|L\}\Pi_r(\alpha_0)}{\Pi_1(\alpha)} - E\{1(R=r)|L\}\right)}_{=0} E\left[U(L;\beta)|L_{(r)}, R=1;\eta^*\right] \\
&= E\left[U(L;\beta_0)\right] = 0
\end{aligned}
$$

By the same token, under $\mathcal{M}_L$, $\widetilde{\alpha} \to \alpha^*$ and $\widetilde{\eta} \to \eta_0$ in probability, then $\mathbb{P}_n V(\beta_0, \widetilde{\alpha}, \widetilde{\eta}) \to E\{V(\beta_0, \alpha^*, \eta_0)\}$. Next we show that $E\{V(\beta_0, \alpha^*, \eta_0)\} = 0$. Note that for all α

$$
\begin{aligned}
\frac{1}{\Pi_1(\alpha)} &= 1 + \sum_{r \neq 1} \frac{\Pi_r(\alpha)}{\Pi_1(\alpha)} \\
&= 1 + \sum_{r \neq 1} \mathrm{Odds}_r\left(L_{(r)};\alpha\right).
\end{aligned}
$$

Then we have that

$$
\begin{aligned}
E\left(V\left(\beta,\alpha_0,\eta^*\right)\right) = & E\left\{\frac{1\left(R=1\right)}{\Pi_1\left(\alpha^*\right)}\left\{U(L;\beta_0)-\sum_{r\neq 1}\Pi_r\left(\alpha^*\right)E\left[U(L;\beta)|L_{(r)},R=1;\eta_0\right]\right\}\right. \\
& \left.+\sum_{r\neq 1}1\left(R=r\right)E\left[U(L;\beta)|L_{(r)},R=1;\eta_0\right]\right\} \\
= & E\left\{1\left(R=1\right)\left\{\frac{U(L;\beta_0)}{\Pi_1\left(\alpha^*\right)}-\sum_{r\neq 1}\frac{\Pi_r\left(\alpha^*\right)}{\Pi_1\left(\alpha^*\right)}E\left[U(L;\beta)|L_{(r)},R=1;\eta_0\right]\right\}\right. \\
& \left.+\sum_{r\neq 1}1\left(R=r\right)E\left[U(L;\beta)|L_{(r)},R=1;\eta_0\right]\right\} \\
= & E\left\{\underbrace{\sum_{r\neq 1}\text{Odds}_r\left(L_{(r)};\alpha^*\right)\left(E\left[U(L;\beta_0)|R=1,L_{(r)}\right]-E\left[U(L;\beta)|L_{(r)},R=1;\eta_0\right]\right)}_{=0}\right. \\
& \left.1\left(R=1\right)U(L;\beta_0)+\sum_{r\neq 1}1\left(R=r\right)E\left[U(L;\beta)|L_{(r)},R=1;\eta_0\right]\right\} \\
= & E\left\{1\left(R=1\right)U(L;\beta_0)+\sum_{r\neq 1}1\left(R=r\right)E\left[U(L;\beta)|L_{(r)},R=1;\eta_0\right]\right\} \\
= & E\left\{1\left(R=1\right)U(L;\beta_0)+\sum_{r\neq 1}1\left(R=r\right)E\left[U(L;\beta)|L_{(r)},R=r\right]\right\} \\
= & E\left\{1\left(R=1\right)E[U(L;\beta_0)|R=1]+\sum_{r\neq 1}1\left(R=r\right)E\left[U(L;\beta)|R=r\right]\right\} \\
= & E[U(L;\beta_0)]=0
\end{aligned}
$$

proving the result.

Proof of Corollary 4: $E\left(V\left(\beta,\alpha,\eta\right)\right)$ can be written

$$
\begin{aligned}
&E\left(V\left(\beta,\alpha,\eta\right)\right)\\
&=E\left\{\sum_{r\neq1}\frac{1\left(R=1\right)\Pi_r\left(\alpha\right)}{\Pi_1\left(\alpha\right)}U(L;\beta_0)-\sum_{r\neq1}\frac{1\left(R=1\right)\Pi_r\left(\alpha\right)}{\Pi_1\left(\alpha\right)}E\left[U(L;\beta)|L_{(r)},R=1;\eta\right]\right.\\
&\left.+\sum_{r\neq1}1\left(R=r\right)E\left[U(L;\beta)|L_{(r)},R=1;\eta\right]+1\left(R=1\right)U(L;\beta_0)\right\}\\
&=E\left\{\sum_{r\neq1}\frac{1\left(R=1\right)\Pi_r\left(\alpha\right)}{\Pi_1\left(\alpha\right)}U(L;\beta_0)-\sum_{r\neq1}\frac{1\left(R=1\right)\Pi_r\left(\alpha\right)}{\Pi_1\left(\alpha\right)}E\left[U(L;\beta)|L_{(r)},R=1;\eta\right]\right.\\
&\left.+\sum_{r\neq1}1\left(R=r\right)\left\{E\left[U(L;\beta)|L_{(r)},R=1;\eta\right]-U(L;\beta_0)\right\}+U(L;\beta_0)\right\}\\
&=E\left[\sum_{r\neq1}\left\{1\left(R=1\right)\text{Odds}_r\left(L_{(r)};\alpha\right)-1\left(R=r\right)\right\}\left\{U(L;\beta_0)-E\left[U(L;\beta)|L_{(r)},R=1;\eta\right]\right\}\right]
\end{aligned}
$$

Under $\mathcal{M}_R\left(r\right)$, we have that $\text{Odds}_r\left(L_{(r)};\widetilde{\alpha}\right)\rightarrow\text{Odds}_r\left(L_{(r)};\alpha_0\right)$ in probability, and

$$
\begin{aligned}
&E\left[\left\{1\left(R=1\right)\text{Odds}_r\left(L_{(r)};\alpha_0\right)-1\left(R=r\right)\right\}\left\{U(L;\beta_0)-E\left[U(L;\beta)|L_{(r)},R=1;\eta^*\right]\right\}\right]\\
&=E\left[\left\{1\left(R=1\right)\frac{\Pi_r}{\Pi_1}-1\left(R=r\right)\right\}\left\{U(L;\beta_0)-E\left[U(L;\beta)|L_{(r)},R=1;\eta^*\right]\right\}\right]\\
&=E\left[\left\{\Pi_r-E\left[1\left(R=r\right)|L\right]\right\}\left\{U(L;\beta_0)-E\left[U(L;\beta)|L_{(r)},R=1;\eta^*\right]\right\}\right]\\
&=0
\end{aligned}
$$

Likewise, under $\mathcal{M}_L\left(r\right)$, we have that $E\left[U(L;\beta)|L_{(r)},R=1;\widetilde{\eta}\right]\rightarrow E\left[U(L;\beta)|L_{(r)},R=1;\eta_0\right]$ in probability, and

$$
\begin{aligned}
&E\left[\left\{1\left(R=1\right)\mathrm{Odds}_r\left(L_{(r)};\alpha^*\right)-1\left(R=r\right)\right\}\left\{U(L;\beta_0)-E\left[U(L;\beta)|L_{(r)},R=1;\eta_0\right]\right\}\right]\\
&=E\left[1\left(R=1\right)\mathrm{Odds}_r\left(L_{(r)};\alpha^*\right)\left\{E\left\{U(L;\beta_0)|R=1,L_{(r)}\right\}-E\left[U(L;\beta)|L_{(r)},R=1;\eta_0\right]\right\}\right]\\
&-E\left[\{1\left(R=r\right)\}\left\{E\left\{U(L;\beta_0)|R=r,L_{(r)}\right\}-E\left[U(L;\beta)|L_{(r)},R=1;\eta_0\right]\right\}\right]\\
&=-E\left[\{1\left(R=r\right)\}\left\{E\left\{U(L;\beta_0)|R=1,L_{(r)}\right\}-E\left[U(L;\beta)|L_{(r)},R=1;\eta_0\right]\right\}\right]\\
&=0
\end{aligned}
$$

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 β is 0.634, and the sample size is 2000.

	`bth`*	`nrm`	`ccm`	`bad`
Bias(SE)				
IPW	-0.004(0.002)	-0.004(0.002)	-0.641(0.012)	-0.641(0.012)
PM	-0.002(0.001)	-0.367(0.002)	-0.002(0.001)	-0.367(0.002)
DR	-0.002(0.002)	-0.006(0.002)	-0.002(0.002)	-0.371(0.003)
RMSE				
IPW	0.072	0.072	0.748	0.748
PM	0.046	0.373	0.046	0.373
DR	0.048	0.057	0.057	0.385

*: `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 β to corresponding Monte Carlo standard deviations .

References

[1] Chen, H. Y. (2007). A semiparametric odds ratio model for measuring association. Biometrics **63**, 413-421.

[2] Robins JM, Rotnitzky A, Scharfstein D. (1999). Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models. In: Statistical Models in Epidemiology: The Environment and Clinical Trials. Halloran, M.E. and Berry, D., eds. IMA Volume 116, NY: Springer-Verlag, pp. 1-92.

[3] Tchetgen Tchetgen, E. J., Robins, J. M., & Rotnitzky, A. (2010). On doubly robust estimation in a semiparametric odds ratio model. Biometrika **97**, 171-180.