Web-based Supplementary Materials for 'Identifiability and Estimation of Causal Effects in Randomized Trials with Noncompliance and Completely Non-ignorable Missing-Data' By Chen et al.

Appendix Web Appendix A: Proof of Theorem 1

Identifiability of ξ, ω_a and ω_n is immediate from randomization of Z and the monotonicity assumption, that is, $\xi = P(Z = 1)$, $\omega_a = P(U = a) = P(D = 1|Z = 0)$ and $\omega_n = P(U = n) = P(D = 0|Z = 1)$. We next show that δ_{yzu} are the functions of the distributions of observed variables. Under the Assumption 3, we obtain that $\delta_{y1n} = P(Y = y, R = 1|Z = 1, U = n) = \frac{P(Y=y,R=1,Z=1,D=0)}{P(Z=1,D=0)}$ and that $\delta_{y0a} = \frac{P(Y=y,R=1,Z=0,D=1)}{P(Z=0,D=1)}$.

For δ_{y1c}, we have $\delta_{y1c} = \frac{P(Y=y,R=1,Z=1,U=c)}{P(Z=1,U=c)} = \frac{P(Y=y,R=1,Z=1,D=1)-P(Y=y,R=1,D=1,Z=1,U=a)}{P(Z=1,D=1)-P(D=1,Z=1,U=a)}$. Under the monotonicity and randomization assumptions, $P(D = 1, Z = 1, U = a)$ in the denominator can be rewritten as $P(Z = 1)P(D = 1, U = a|Z = 1) = P(Z = 1)P(U = a|Z = 1) = P(Z = 1)P(U = a|Z = 0) = P(Z = 1)P(D = 1, U = a|Z = 0) = P(Z = 1)P(D = 1|Z = 0)$. On the other hand, from the numerator we have that $P(Y = y, R = 1, D = 1, Z = 1, U = a) = P(R = 1|Y = y, D = 1, Z = 1, U = a)P(Y = y|D = 1, Z = 1, U = a)P(D = 1, Z = 1, U = a)$, where $P(R = 1|Y = y, D = 1, Z = 1, U = a) = P(R = 1|Y = y) = P(R = 1|Y = y, D = 1, Z = 0, U = a)$ because of Assumption 6, $P(Y = y|D = 1, Z = 1, U = a) = P(Y = y|D = 1, Z = 0, U = a)$ due to the exclusion restriction and $P(D = 1, Z = 1, U = a) = P(D = 1, U = a|Z = 0)P(Z = 1)$ by the forward proof. So $P(Y = y, R = 1, D = 1, Z = 1, U = a) = P(R = 1|Y = y, D = 1, Z = 0, U = a)P(Y = y|D = 1, Z = 0, U = a)P(D = 1, U = a|Z = 0)P(Z = 1) = P(Y = y, R = 1, D = 1, U = a|Z = 0)P(Z = 1)$. Hence, we obtain that $\delta_{y1c} = \frac{P(Y=y,R=1,Z=1,D=1)-P(Y=y,R=1,D=1|Z=0)P(Z=1)}{P(Z=1,D=1)-P(D=1|Z=0)P(Z=1)}$.

Similarly, we can show that $\delta_{y0c} = \frac{P(Y=y,R=1,Z=0,D=0)-P(Y=y,R=1,D=0|Z=1)P(Z=0)}{P(Z=0,D=0)-P(D=0|Z=1)P(Z=0)}$. Hence, we have shown that δ_{yzu}'s are identifiable.

Next we will show that ρ_y's are identifiable. Let us define the matrix $\underline{\boldsymbol{\Delta}}^1$ as follows:

$$\underline{\boldsymbol{\Delta}}^1 = \begin{pmatrix} \delta_{01n} & \delta_{00a} & \delta_{01c} & \delta_{00c} \\ \delta_{11n} & \delta_{10a} & \delta_{11c} & \delta_{10c} \end{pmatrix}^T .$$

Because $\theta_{0zu} + \theta_{1zu} = 1$ and from $\delta_{yzu} = \rho_y \theta_{yzu}$, we obtain the following equations:

$$\underline{\boldsymbol{\Delta}}^1 \begin{pmatrix} 1/\rho_0 \\ 1/\rho_1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}^T . \tag{A.1}$$

Below we show that $\underline{\boldsymbol{\Delta}}^1$ has rank 2. Suppose that $\underline{\boldsymbol{\Delta}}^1$ does not have full column rank. Then we have $\frac{\delta_{01n}}{\delta_{11n}} = \frac{\delta_{00a}}{\delta_{10a}} = \frac{\delta_{01c}}{\delta_{11c}} = \frac{\delta_{00c}}{\delta_{10c}}$, which implies $\frac{\theta_{01n}}{\theta_{11n}} = \frac{\theta_{00a}}{\theta_{10a}} = \frac{\theta_{01c}}{\theta_{11c}} = \frac{\theta_{00c}}{\theta_{10c}}$ since $\delta_{yzu} = \rho_y \theta_{yzu}$. Thus we obtain that $\theta_{10a} = \theta_{11n} = \theta_{11c} = \theta_{10c}$, which implies that Y is independent of Z given U and is also independent of U given Z. This contradicts the condition of Theorem 1. Therefore, we have shown that ρ_y's are identifiable. Finally, the parameters θ_{yzu} can be identified from equations: $\theta_{10n} = \theta_{11n} = \delta_{11n}/\rho_1$, $\theta_{11a} = \theta_{10a} = \delta_{10a}/\rho_1$, $\theta_{11c} = \delta_{11c}/\rho_1$ and $\theta_{10c} = \delta_{10c}/\rho_1$.

Appendix Web Appendix B: Moment Estimators under the Model specified in Theorem 1

COROLLARY 1: *Under the assumptions of Theorem 1, the moment estimator, $\hat{\theta}$, of the parameter vector θ is given as follows: for $y = 0$ and 1,* $\hat{\delta}_{y1n} = \frac{N_{y110}}{N_{++10}}, \hat{\delta}_{y0a} = \frac{N_{y101}}{N_{++01}}$,

$$\hat{\delta}_{y1c} = \frac{N_{y111} - N_{y101}\frac{N_{++1+}}{N_{++0+}}}{N_{++11} - N_{++01}\frac{N_{++1+}}{N_{++0+}}}, \; \hat{\delta}_{y0c} = \frac{N_{y100} - N_{y110}\frac{N_{++0+}}{N_{++1+}}}{N_{++00} - N_{++10}\frac{N_{++0+}}{N_{++1+}}}.$$

$\hat{\xi} = \frac{N_{++1+}}{N}$, $\hat{\omega}_a = \frac{N_{++01}}{N_{++0+}}$, $\hat{\omega}_n = \frac{N_{++10}}{N_{++1+}}, \hat{\theta}_{10a} = \frac{\hat{\delta}_{10a}}{\hat{\rho}_1}$, $\hat{\theta}_{11n} = \frac{\hat{\delta}_{11n}}{\hat{\rho}_1}$, $\hat{\theta}_{11c} = \frac{\hat{\delta}_{11c}}{\hat{\rho}_1}$, $\hat{\theta}_{10c} = \frac{\hat{\delta}_{10c}}{\hat{\rho}_1}$. *Here $\hat{\rho}_0$ and $\hat{\rho}_1$ are computed as follows:*

(1) if $P(Y = 1|Z = 1, U = n) \neq P(Y = 1|Z = 1, U = a)$ (i.e. $\theta_{11n} \neq \theta_{10a}$), then $\hat{\rho}_0 = \frac{\hat{\delta}_{01n}\hat{\delta}_{10a} - \hat{\delta}_{11n}\hat{\delta}_{00a}}{\hat{\delta}_{10a} - \hat{\delta}_{11n}}$, $\hat{\rho}_1 = \frac{\hat{\delta}_{01n}\hat{\delta}_{10a} - \hat{\delta}_{11n}\hat{\delta}_{00a}}{\hat{\delta}_{01n} - \hat{\delta}_{00a}}$;

(2) if $P(Y = 1|Z = 1, U = n) = P(Y = 1|Z = 1, U = a)$(i.e. $\theta_{11n} = \theta_{10a}$) and $P(Y = 1|Z =$

$1, U = c) \neq P(Y = 1|Z = 0, U = c)$ *(i.e.* $\theta_{11c} \neq \theta_{10c}$*), then* $\hat{\rho}_0 = \frac{\hat{\delta}_{01c}\hat{\delta}_{10c} - \hat{\delta}_{11c}\hat{\delta}_{00c}}{\hat{\delta}_{10c} - \hat{\delta}_{11c}}$, $\hat{\rho}_1 = \frac{\hat{\delta}_{01c}\hat{\delta}_{10c} - \hat{\delta}_{11c}\hat{\delta}_{00c}}{\hat{\delta}_{01c} - \hat{\delta}_{00c}}$;

(3) if $P(Y = 1|Z = 1, U = n) = P(Y = 1|Z = 1, U = a)$ *(i.e.* $\theta_{11n} = \theta_{10a}$*) and* $P(Y = 1|Z = 1, U = c) = P(Y = 1|Z = 0, U = c)$ *(i.e.* $\theta_{11c} = \theta_{10c}$*), then* $\hat{\rho}_0 = \frac{\hat{\delta}_{01n}\hat{\delta}_{10c} - \hat{\delta}_{11n}\hat{\delta}_{00c}}{\hat{\delta}_{10c} - \hat{\delta}_{11n}}$, $\hat{\rho}_1 = \frac{\hat{\delta}_{01n}\hat{\delta}_{10c} - \hat{\delta}_{11n}\hat{\delta}_{00c}}{\hat{\delta}_{01n} - \hat{\delta}_{00c}}$.

Appendix Web Appendix C: The EM algorithm under the assumptions in Theorem 1

Let $\theta^{(k)} = (\omega_a^{(k)}, \omega_n^{(k)}, \theta_{10a}^{(k)}, \theta_{11n}^{(k)}, \theta_{11c}^{(k)}, \theta_{10c}^{(k)}, \rho_0^{(k)}, \rho_1^{(k)})$ be the estimate of θ after the kth iteration in the EM algorithm. Define $\pi_0^{(k)} = (\omega_n^{(k)}\theta_{10n}^{(k)} + (1-\omega_n^{(k)}-\omega_a^{(k)})\theta_{10c}^{(k)})(1-\rho_1^{(k)}) + (\omega_n^{(k)}(1-\theta_{10n}^{(k)}) + (1-\omega_n^{(k)}-\omega_a^{(k)})(1-\theta_{10c}^{(k)}))(1-\rho_0^{(k)})$, $\pi_1^{(k)} = (\omega_a^{(k)}\theta_{11a}^{(k)} + (1-\omega_n^{(k)}-\omega_a^{(k)})\theta_{11c}^{(k)})(1-\rho_1^{(k)}) + (\omega_a^{(k)}(1-\theta_{11a}^{(k)}) + (1-\omega_n^{(k)}-\omega_a^{(k)})(1-\theta_{11c}^{(k)}))(1-\rho_0^{(k)})$, and $n_{yrzu} = \#of(Y = y, R = r, Z = z, U = t)$. Then the next iteration estimate $\theta^{(k+1)}$ of θ in the EM algorithm is given as follow:

$n_{110n}^{(k+1)} = N_{1100}\frac{\omega_n^{(k)}\theta_{10n}^{(k)}\rho_1^{(k)}}{\omega_n^{(k)}\theta_{10n}^{(k)}\rho_1^{(k)} + (1-\omega_n^{(k)}-\omega_a^{(k)})\theta_{10c}^{(k)}\rho_1^{(k)}}, n_{110c}^{(k+1)} = N_{1100} - n_{110n}^{(k+1)}$,

$n_{010n}^{(k+1)} = N_{0100}\frac{\omega_n^{(k)}(1-\theta_{10n}^{(k)})\rho_0^{(k)}}{\omega_n^{(k)}(1-\theta_{10n}^{(k)})\rho_0^{(k)} + (1-\omega_n^{(k)}-\omega_a^{(k)})(1-\theta_{10c}^{(k)})\rho_0^{(k)}}, n_{010c}^{(k+1)} = N_{0100} - n_{010n}^{(k+1)}$,

$n_{111n}^{(k+1)} = N_{1110}, n_{011n}^{(k+1)} = N_{0110}, n_{110a}^{(k+1)} = N_{1101}, n_{010a}^{(k+1)} = N_{0101}$,

$n_{111a}^{(k+1)} = N_{1111}\frac{\omega_a^{(k)}\theta_{11a}^{(k)}\rho_1^{(k)}}{\omega_a^{(k)}\theta_{11a}^{(k)}\rho_1^{(k)} + (1-\omega_n^{(k)}-\omega_a^{(k)})\theta_{11c}^{(k)}\rho_1^{(k)}}, n_{111c}^{(k+1)} = N_{1111} - n_{111a}^{(k+1)}$,

$n_{011a}^{(k+1)} = N_{0111}\frac{\omega_a^{(k)}(1-\theta_{11a}^{(k)})\rho_0^{(k)}}{\omega_a^{(k)}(1-\theta_{11a}^{(k)})\rho_0^{(k)} + (1-\omega_n^{(k)}-\omega_a^{(k)})(1-\theta_{11c}^{(k)})\rho_0^{(k)}}, n_{011c}^{(k+1)} = N_{0111} - n_{011a}^{(k+1)}$,

$n_{100n}^{(k+1)} = N_{+000}\frac{\omega_n^{(k)}\theta_{10n}^{(k)}(1-\rho_1^{(k)})}{\pi_0^{(k)}}, n_{100c}^{(k+1)} = N_{+000}\frac{(1-\omega_n^{(k)}-\omega_a^{(k)})\theta_{10c}^{(k)}(1-\rho_1^{(k)})}{\pi_0^{(k)}}$,

$n_{000n}^{(k+1)} = N_{+000}\frac{\omega_n^{(k)}(1-\theta_{10n}^{(k)})(1-\rho_0^{(k)})}{\pi_0^{(k)}}, n_{000c}^{(k+1)} = N_{+000}\frac{(1-\omega_n^{(k)}-\omega_a^{(k)})(1-\theta_{10c}^{(k)})(1-\rho_0^{(k)})}{\pi_0^{(k)}}$,

$n_{101n}^{(k+1)} = N_{+010}\frac{\theta_{11n}^{(k)}(1-\rho_1^{(k)})}{\theta_{11n}^{(k)}(1-\rho_1^{(k)}) + (1-\theta_{11n}^{(k)})(1-\rho_0^{(k)})}, n_{001n}^{(k+1)} = N_{+010} - n_{101n}^{(k+1)}$,

$n_{100a}^{(k+1)} = N_{+001}\frac{\theta_{10a}^{(k)}(1-\rho_1^{(k)})}{\theta_{10a}^{(k)}(1-\rho_1^{(k)}) + (1-\theta_{10a}^{(k)})(1-\rho_0^{(k)})}, n_{000a}^{(k+1)} = N_{+001} - n_{100a}^{(k+1)}$,

$n_{101a}^{(k+1)} = N_{+011}\frac{\omega_a^{(k)}\theta_{11a}^{(k)}(1-\rho_1^{(k)})}{\pi_1^{(k)}}, n_{101c}^{(k+1)} = N_{+011}\frac{(1-\omega_n^{(k)}-\omega_a^{(k)})\theta_{11c}^{(k)}(1-\rho_1^{(k)})}{\pi_1^{(k)}}$,

$n_{001a}^{(k+1)} = N_{+011}\frac{\omega_a^{(k)}(1-\theta_{11a}^{(k)})(1-\rho_0^{(k)})}{\pi_1^{(k)}}, n_{001c}^{(k+1)} = N_{+011}\frac{(1-\omega_n^{(k)}-\omega_a^{(k)})(1-\theta_{11c}^{(k)})(1-\rho_0^{(k)})}{\pi_1^{(k)}}$,

$\omega_a^{(k+1)} = \frac{n_{000a}^{(k+1)} + n_{001a}^{(k+1)} + n_{010a}^{(k+1)} + n_{011a}^{(k+1)} + n_{100a}^{(k+1)} + n_{101a}^{(k+1)} + n_{110a}^{(k+1)} + n_{111a}^{(k+1)}}{N}$,

$\omega_n^{(k+1)} = \frac{n_{000n}^{(k+1)}+n_{001n}^{(k+1)}+n_{010n}^{(k+1)}+n_{011n}^{(k+1)}+n_{100n}^{(k+1)}+n_{101n}^{(k+1)}+n_{110n}^{(k+1)}+n_{111n}^{(k+1)}}{N}$,

$\theta_{10a}^{(k+1)} = \frac{n_{100a}^{(k+1)}+n_{101a}^{(k+1)}+n_{110a}^{(k+1)}+n_{111a}^{(k+1)}}{n_{000a}^{(k+1)}+n_{001a}^{(k+1)}+n_{010a}^{(k+1)}+n_{011a}^{(k+1)}+n_{100a}^{(k+1)}+n_{101a}^{(k+1)}+n_{110a}^{(k+1)}+n_{111a}^{(k+1)}}$,

$\theta_{11n}^{(k+1)} = \frac{n_{100n}^{(k+1)}+n_{101n}^{(k+1)}+n_{110n}^{(k+1)}+n_{111n}^{(k+1)}}{n_{000n}^{(k+1)}+n_{001n}^{(k+1)}+n_{010n}^{(k+1)}+n_{011n}^{(k+1)}+n_{100n}^{(k+1)}+n_{101n}^{(k+1)}+n_{110n}^{(k+1)}+n_{111n}^{(k+1)}}$,

$\theta_{11c}^{(k+1)} = \frac{n_{101c}^{(k+1)}+n_{111c}^{(k+1)}}{n_{101c}^{(k+1)}+n_{111c}^{(k+1)}+n_{001c}^{(k+1)}+n_{011c}^{(k+1)}}, \theta_{10c}^{(k+1)} = \frac{n_{100c}^{(k+1)}+n_{110c}^{(k+1)}}{n_{100c}^{(k+1)}+n_{110c}^{(k+1)}+n_{000c}^{(k+1)}+n_{010c}^{(k+1)}}$,

$\rho_0^{(k+1)} = \frac{n_{010n}^{(k+1)}+n_{011n}^{(k+1)}+n_{010a}^{(k+1)}+n_{011a}^{(k+1)}+n_{010c}^{(k+1)}+n_{011c}^{(k+1)}}{D_0}$,

$\rho_1^{(k+1)} = \frac{n_{110n}^{(k+1)}+n_{111n}^{(k+1)}+n_{110a}^{(k+1)}+n_{111a}^{(k+1)}+n_{110c}^{(k+1)}+n_{111c}^{(k+1)}}{D_1}$, where $n_{yrzt}^{(k+1)} = E(n_{yrzt} \mid \text{observed data}, \theta = \theta^{(k)})$, $D_0 = n_{010n}^{(k+1)} + n_{011n}^{(k+1)} + n_{010a}^{(k+1)} + n_{011a}^{(k+1)} + n_{010c}^{(k+1)} + n_{011c}^{(k+1)} + n_{000n}^{(k+1)} + n_{001n}^{(k+1)} + n_{000a}^{(k+1)} + n_{001a}^{(k+1)} + n_{000c}^{(k+1)} + n_{001c}^{(k+1)}$, $D_1 = n_{110n}^{(k+1)} + n_{111n}^{(k+1)} + n_{110a}^{(k+1)} + n_{111a}^{(k+1)} + n_{110c}^{(k+1)} + n_{111c}^{(k+1)} + n_{100n}^{(k+1)} + n_{101n}^{(k+1)} + n_{100a}^{(k+1)} + n_{101a}^{(k+1)} + n_{100c}^{(k+1)} + n_{101c}^{(k+1)}$.

Proof. Let $n_{yrzu} = \sum_{i=1}^{N} I_{\{Y_i=y,R_i=r,Z_i=z,U_i=u\}}$. The complete-data likelihood function is given as follows:

$L_c(\theta) = \Pi_{i=1}^{N} P(Z_i)P(U_i)P(D_i|Z_i,U_i)P(Y_i|Z_i,U_i)P(R_i|Y_i)$, where $P(Z_i = 1) = \xi$ can be dropped during the EM steps. Then the complete-data log-likelihood function is given as follows:

$l_c(\theta) = n_{110n}log(\omega_n\theta_{10n}\rho_1) + n_{110c}log((1-\omega_n-\omega_a)\theta_{10c}\rho_1)$

$+n_{010n}log(\omega_n(1-\theta_{10n})\rho_0) + n_{010c}log((1-\omega_n-\omega_a)(1-\theta_{10c})\rho_0)$

$+n_{111n}log(\omega_n\theta_{11n}\rho_1) + n_{011n}log(\omega_n(1-\theta_{11n})\rho_0) + n_{110a}log(\omega_a\theta_{10a}\rho_1)$

$+n_{010a}log(\omega_a(1-\theta_{10a})\rho_0) + n_{111a}log(\omega_a\theta_{11a}\rho_1) + n_{111c}log((1-\omega_n-\omega_a)\theta_{11c}\rho_1)$

$+n_{011a}log(\omega_a(1-\theta_{11a})\rho_0) + n_{011c}log((1-\omega_n-\omega_a)(1-\theta_{11c})\rho_0)$

$+n_{100n}log(\omega_n\theta_{10n}(1-\rho_1)) + n_{100c}log((1-\omega_n-\omega_a)\theta_{10c}(1-\rho_1))$

$+n_{000n}log(\omega_n(1-\theta_{10n})(1-\rho_0)) + n_{000c}log((1-\omega_n-\omega_a)(1-\theta_{10c})(1-\rho_0))$

$+n_{101n}log(\omega_n\theta_{11n}(1-\rho_1)) + n_{001n}log(\omega_n(1-\theta_{11n})(1-\rho_0))$

$+n_{100a}log(\omega_a\theta_{10a}(1-\rho_1)) + n_{000a}log(\omega_a(1-\theta_{10a})(1-\rho_0))$

$+n_{101a}log(\omega_a\theta_{11a}(1-\rho_1)) + n_{101c}log((1-\omega_n-\omega_a)\theta_{11c}(1-\rho_1))$

$+n_{001a}log(\omega_a(1-\theta_{11a})(1-\rho_0)) + n_{001c}log((1-\omega_n-\omega_a)(1-\theta_{11c})(1-\rho_0))$. In the E step, we take the expectation of the complete-data log-likelihood, given the observed data and the previous parameter estimate $\theta = \theta^{(k)}$, that is $n_{yrzu}^{(k+1)} = E[n_{yrzu}|observed - data, \theta^{(k)}]$. In the M step, we can solve following functions and get the estimates.

$$\frac{\partial l}{\partial \omega_a} = \frac{n_{000a}^{(k+1)}+n_{001a}^{(k+1)}+n_{010a}^{(k+1)}+n_{011a}^{(k+1)}+n_{100a}^{(k+1)}+n_{101a}^{(k+1)}+n_{110a}^{(k+1)}+n_{111a}^{(k+1)}}{\omega_a}$$
$$-\frac{n_{000c}^{(k+1)}+n_{001c}^{(k+1)}+n_{010c}^{(k+1)}+n_{011c}^{(k+1)}+n_{100c}^{(k+1)}+n_{101c}^{(k+1)}+n_{110c}^{(k+1)}+n_{111c}^{(k+1)}}{1-\omega_n-\omega_a},$$
$$\frac{\partial l}{\partial \omega_n} = \frac{n_{000n}^{(k+1)}+n_{001n}^{(k+1)}+n_{010n}^{(k+1)}+n_{011n}^{(k+1)}+n_{100n}^{(k+1)}+n_{101n}^{(k+1)}+n_{110n}^{(k+1)}+n_{111n}^{(k+1)}}{\omega_n}$$
$$-\frac{n_{000c}^{(k+1)}+n_{001c}^{(k+1)}+n_{010c}^{(k+1)}+n_{011c}^{(k+1)}+n_{100c}^{(k+1)}+n_{101c}^{(k+1)}+n_{110c}^{(k+1)}+n_{111c}^{(k+1)}}{1-\omega_n-\omega_a},$$
$$\frac{\partial l}{\partial \theta_{10a}} = \frac{n_{100a}^{(k+1)}+n_{101a}^{(k+1)}+n_{110a}^{(k+1)}+n_{111a}^{(k+1)}}{\theta_{10a}}$$
$$-\frac{n_{000a}^{(k+1)}+n_{001a}^{(k+1)}+n_{010a}^{(k+1)}+n_{011a}^{(k+1)}}{1-\theta_{10a}},$$
$$\frac{\partial l}{\partial \theta_{11n}} = \frac{n_{100n}^{(k+1)}+n_{101n}^{(k+1)}+n_{110n}^{(k+1)}+n_{111n}^{(k+1)}}{\theta_{11n}}$$
$$-\frac{n_{000n}^{(k+1)}+n_{001n}^{(k+1)}+n_{010n}^{(k+1)}+n_{011n}^{(k+1)}}{1-\theta_{11n}},$$
$$\frac{\partial l}{\partial \theta_{11c}} = \frac{n_{101c}^{(k+1)}+n_{111c}^{(k+1)}}{\theta_{11c}} - \frac{n_{001c}^{(k+1)}+n_{011c}^{(k+1)}}{1-\theta_{11c}}, \quad \frac{\partial l}{\partial \theta_{10c}} = \frac{n_{100c}^{(k+1)}+n_{110c}^{(k+1)}}{\theta_{10c}} - \frac{n_{000c}^{(k+1)}+n_{010c}^{(k+1)}}{1-\theta_{10c}},$$
$$\frac{\partial l}{\partial \rho_0} = \frac{n_{010n}^{(k+1)}+n_{011n}^{(k+1)}+n_{010a}^{(k+1)}+n_{011a}^{(k+1)}+n_{010c}^{(k+1)}+n_{011c}^{(k+1)}}{\rho_0}$$
$$-\frac{n_{000n}^{(k+1)}+n_{001n}^{(k+1)}+n_{000a}^{(k+1)}+n_{001a}^{(k+1)}+n_{000c}^{(k+1)}+n_{001c}^{(k+1)}}{1-\rho_0},$$
$$\frac{\partial l}{\partial \rho_1} = \frac{n_{110n}^{(k+1)}+n_{111n}^{(k+1)}+n_{110a}^{(k+1)}+n_{111a}^{(k+1)}+n_{110c}^{(k+1)}+n_{111c}^{(k+1)}}{\rho_1}$$
$$-\frac{n_{100n}^{(k+1)}+n_{101n}^{(k+1)}+n_{100a}^{(k+1)}+n_{101a}^{(k+1)}+n_{100c}^{(k+1)}+n_{101c}^{(k+1)}}{1-\rho_1}.$$

Appendix Web Appendix B: Proof of Theorem 2

The joint distribution can be factorized as $P(Z, U, D, X, Y, R) = P(R|Z, U, D, X, Y)\ P(Y|Z, U, D, X)$ $P(D|Z, U, X)P(U, X|Z)P(Z)$. Since Z is randomized, $P(U, X|Z) = P(U, X)$. Because D is determined by (Z, U), we obtain that $P(D|Z, U, X) = P(D|Z, U)$ and $P(Y|Z, U, D, X) = P(Y|Z, U, X)$. From the randomization assumption and Assumption 7, we obtain that $P(R|Z, U, D, X, Y) = P(R|Y, Z)$. So we can rewrite the joint distribution as

$$P(Z, U, D, X, Y, R) = P(Z)P(U, X)P(D|Z, U)P(Y|Z, U, X)P(R|Y, Z).$$

To identify $P(U,X)$, we first note that from independence of Z and (U,X), the definition of U, and Assumption 3, we obtain that $P(U=a|X=x)=P(U=a|Z=0,X=x)=P(D=1|Z=0,X=x)$, $P(U=n|X=x)=P(U=n|Z=1,X=x)=P(D=0|Z=1,X=x)$ and then $P(U=c|X=x)=1-\{P(U=n|X=x)+P(U=a|X=x)\}$. Since $P(X)$ is identifiable, $P(U,X)$ is identifiable; since Z and U determine D, $P(D|Z=z,U=u)$ is known for all z and u.

Below we show that ρ_{01}, ρ_{11}, ρ_{00}, ρ_{10}, θ_{11nx}, θ_{10ax}, θ_{11cx} and θ_{10cx} are identifiable conditionally on $X=x$. Let us define the following matrices:

$$\underline{\boldsymbol{\Delta}}^2_{zu} = \begin{pmatrix} \delta_{0zux} & \delta_{1zux} \\ \delta_{0zux'} & \delta_{1zux'} \end{pmatrix}$$

Since

$$\delta_{yzux} = P(R=1|Y=y,Z=z)P(Y=y|Z=z,U=u,X=x) = \rho_{yz}\theta_{yzux}, \quad \text{(A.2)}$$

using the same idea as in the proof of Theorem 1, we obtain the following equations:

$$\underline{\boldsymbol{\Delta}}^2_{1n} \begin{pmatrix} 1/\rho_{01} \\ 1/\rho_{11} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{(A.3)}$$

and

$$\underline{\boldsymbol{\Delta}}^2_{0a} \begin{pmatrix} 1/\rho_{00} \\ 1/\rho_{10} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad \text{(A.4)}$$

Under Assumption 3, the elements in the matrices $\underline{\boldsymbol{\Delta}}^2_{1n}$ and $\underline{\boldsymbol{\Delta}}^2_{0a}$ can be expressed by the distributions of observed variables, respectively, as follows:

$$\delta_{y1nx} = \frac{P(Y=y,R=1,Z=1,D=0,X=x)}{P(Z=1,D=0,X=x)} \quad \text{(A.5)}$$

and

$$\delta_{y0ax} = \frac{P(Y=y,R=1,Z=0,D=1,X=x)}{P(Z=0,D=1,X=x)}. \quad \text{(A.6)}$$

Suppose that $\underline{\boldsymbol{\Delta}}^2_{1n}$ in (A.3) is not full rank for all $x \neq x'$. Then it is immediate that Y is independent of X given $U=n$ and $Z=z$, which contradicts the assumptions in Theorem 2.

Thus there exists at least one pair of x and x' so that the matrix $\underline{\mathbf{\Delta}}^2_{1n}$ is full rank, and then ρ_{01} and ρ_{11} can be solved from (A.3). Similarly, we can show that ρ_{00} and ρ_{10} are identifiable from (A.4).

From (A.2), (A.5) and (A.6), we can identify θ_{y1nx} and θ_{y0ax}. Similarly we can identify θ_{y1cx} and θ_{y0cx} from the equations $\theta_{y1cx} = \frac{\delta_{y1cx}}{\rho_{y1}} = \frac{P(y,R=1,Z=1,D=1,x)/(\xi\rho_{y1}) - P(y,R=1,Z=0,D=1,x)/\{(1-\xi)\rho_{y0}\}}{P(Z=1,D=1,x)/\xi - P(D=1,Z=0,x)/(1-\xi)}$ and $\theta_{y0cx} = \frac{P(y,R=1,Z=0,D=0,x)/\{(1-\xi)\rho_{y0}\} - P(y,R=1,Z=1,D=0,x)/(\xi\rho_{y1})}{P(Z=0,D=0,x)/(1-\xi) - P(D=0,Z=1,x)/\xi}$.

Appendix Web Appendix C: Proof of Theorem 3

Similar to the proof of Theorem 2, we have $P(Z,U,D,X,Y,R) = P(Z)P(U,X)P(D|Z,U)$ $P(Y|Z,U,X)P(R|Y,Z,U)$. Hence, we can identify δ_{y1nx} and δ_{y0ax} from (A.5) and (A.6) respectively. We can also identify δ_{y1cx} and δ_{y0cx} by $\delta_{y1cx} = \frac{P(y,R=1,Z=1,D=1,x) - P(y,R=1,D=1,Z=0,x)\xi/(1-\xi)}{P(Z=1,D=1,x) - P(D=1,Z=0,x)P(Z=1)/P(Z=0)}$ and $\delta_{y0cx} = \frac{P(y,R=1,Z=0,D=0,x) - P(y,R=1,D=0,Z=1,x)(1-\xi)/\xi}{P(Z=0,D=0,x) - P(D=0,Z=1,x)P(Z=0)/P(Z=1)}$.

Next we show that we can identify ρ_{yzu}. We first put δ_{yzux}'s into the following matrices:

$$\underline{\mathbf{\Delta}}^3_{zu} = \begin{pmatrix} \delta_{0zux} & \delta_{1zux} \\ \delta_{0zux'} & \delta_{1zux'} \end{pmatrix},$$

for $(z,u) = (1,n)$, $(1,c)$, $(0,a)$ and $(0,c)$. Because R is independent of X given (Y,Z,U), we obtain that $\delta_{yzux} = P(R=1|Y=y,Z=z,U=u)P(Y=y|Z=z,U=u,X=x) = \rho_{yzu}\theta_{yzux}$. Hence, we can obtain the following equations:

$$\underline{\mathbf{\Delta}}^3_{zu} \begin{pmatrix} 1/\rho_{0zu} \\ 1/\rho_{1zu} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Using the same argument as in the proof of Theorem 2, we can show that all four $\underline{\mathbf{\Delta}}^3_{zu}$ matrices have full ranks under the assumptions in Theorem 3. Therefore, we can identify ρ_{yzu}'s. Finally, because $\delta_{yzux} = P(R=1|Y=y,Z=z,U=u)P(Y=y|Z=z,U=u,X=x) = \rho_{yzu}\theta_{yzux}$, we can identify $\theta_{10ax}, \theta_{11nx}, \theta_{11cx}$ and θ_{10cx} respectively.