Supplemental Appendix:

Efficiency in estimation under monotonic attrition

Jean-Louis Barnwell — Analysis Group

Saraswata Chaudhuri[19] — McGill University

Table of Contents

The numbering of the equations, corollaries, lemmas and propositions in this supplemental appendix is consistent with the main text of our paper.

[19]Corresponding author. Department of Economics, McGill University and Cireq, Montreal, Canada. Email: saraswata.chaudhuri@mcgill.ca.

A Supplemental Appendix A: Proofs

A.1 Auxiliary lemmas:

Lemma 8 *The MAR condition in* (1) *implies and is implied by the following condition:*

$$P(C = r|T_R) = P(C = r|T_r) \text{ for } r = 1, \ldots, R-1. \tag{24}$$

Proof of Lemma 8: First we show that if (1) holds then (24) also holds. Take any $r = 1, \ldots, R-1$ and note that:

$$\begin{aligned}
P(C = r|T_R) &= P(C = r|T_R, C \geq r) \prod_{k=1}^{r-1} [1 - P(C = k|T_R, C \geq k)] \\
&= P(C = r|T_r, C \geq r) \prod_{k=1}^{r-1} [1 - P(C = k|T_k, C \geq k)] \quad \text{[by (1)]} \\
&= P(C = r|T_r, C \geq r) \prod_{k=1}^{r-1} [1 - P(C = k|T_r, C \geq k)] \quad \text{[by (1)]} \\
&= P(C = r|T_r).
\end{aligned}$$

Now we show that if (24) holds then (1) also holds. Take any $r = 1, \ldots, R-1$ and note that:

$$\begin{aligned}
P(C = r|T_R, C \geq r) &= \frac{P(C = r|T_R)}{P(C \geq r|T_R)} = \frac{P(C = r|T_R)}{1 - P(C \leq r-1|T_R)} = \frac{P(C = r|T_R)}{1 - \sum_{j=1}^{r-1} P(C = j|T_R)} \\
&= \frac{P(C = r|T_r)}{1 - \sum_{j=1}^{r-1} P(C = j|T_j)} = \frac{P(C = r|T_r)}{1 - \sum_{j=1}^{r-1} P(C = j|T_r)} \\
&= \frac{P(C = r|T_r)}{P(C \geq r|T_r)} = P(C = r|T_r, C \geq r)
\end{aligned}$$

where the fourth and fifth equalities follow by (24). ■

Lemma 9 *The MAR condition in* (1) *implies that:*

$$P(C \geq r|T_j) = P(C \geq r|T_{r-1}) \text{ for } r = 1, \ldots, R-1 \text{ and } j = r, \ldots, R.$$

Proof of Lemma 9: Lemma 8 shows that (1) implies (24). Now, take $r = 1, \ldots, R-1$ and $j = r, \ldots, R$ and note that:

$$P(C \geq r|T_j) = 1 - \sum_{k=1}^{r-1} P(C = k|T_j) = 1 - \sum_{k=1}^{r-1} P(C = k|T_k) = 1 - \sum_{k=1}^{r-1} P(C = k|T_{r-1}) = P(C \geq r|T_{r-1})$$

where the second and third equalities follow by (24). ■

Remarks:

1. Lemma 9 implies that if $R = 2$ then $P(C = 2|T_2) = P(C = 2|T_1)$. This is the familiar form in which the MAR assumption is generally found in the econometrics literature where the focus has typically been on the case of $R = 2$.

2. We introduced the notation in the above two lemmas for brevity of expression in the proofs in this appendix. The original notation with the conditional hazards is very transparent in terms of accounting for the observability of the conditioning variables (and hence for estimation), and precisely for that reason it leads to longer expressions.

Lemma 10 *Under the conditions of Proposition 2 and using the notation of Section 3.2:*

$$E\left[\varphi_{[a,b]}(O, \beta^0_{[a,b]})S(O)'\right] = E\left[m \left\{ s(Z) + \frac{\dot{P}(a \leq C \leq b|T_b)}{P(a \leq C \leq b|T_b)} \right\}' \middle| a \leq C \leq b \right].$$

Proof of Lemma 10: Note from (3), (4) and (5) that:

$$\begin{aligned}
\varphi_{[a,b]}(O; \beta^0_{[a,b]}) &= \sum_{r=b+1}^{R} \frac{I(C \geq r)}{P(C \geq r|T_{r-1})} \frac{P(a \leq C \leq b|T_b)}{P(a \leq C \leq b)} \left(E[m|T_r] - E[m|T_{r-1}]\right) \\
&\quad + \sum_{r=a+1}^{b} \frac{I(C \geq r)}{P(C \geq r|T_{r-1})} \frac{P(a \leq C \leq r-1|T_{r-1})}{P(a \leq C \leq b)} \left(E[m|T_r] - E[m|T_{r-1}]\right) \\
&\quad + \sum_{r=a}^{b} \frac{I(C = r)}{P(a \leq C \leq b)} E[m|T_r]. \qquad (25)
\end{aligned}$$

This alternative formulation of $\varphi_{[a,b]}(O; \beta^0_{[a,b]})$ without the conditional hazards stated explicitly is not intuitively transparent for actual computational purpose, but will be adopted here since it provides shorter expressions in the proof. Based on (25) we can write $E[\varphi_{[a,b]}(O;\beta^0_{[a,b]})S(O)'] = \sum_{i=1}^{3}\sum_{j=1}^{2} B_{ij}$ where:

$$B_{11} := \sum_{r=b+1}^{R} E\left[\frac{I(C\geq r)}{P(C\geq r|T_{r-1})}\frac{P(a\leq C\leq b|T_b)}{P(a\leq C\leq b)}\left(E[m|T_r]-E[m|T_{r-1}]\right)D'\right],$$

$$B_{12} := \sum_{r=b+1}^{R} E\left[\frac{I(C\geq r)}{P(C\geq r|T_{r-1})}\frac{P(a\leq C\leq b|T_b)}{P(a\leq C\leq b)}\left(E[m|T_r]-E[m|T_{r-1}]\right)\sum_{k=1}^{R} I(C=k)\frac{\dot{P}(C=k|T_k)'}{P(C=k|T_k)}\right],$$

$$B_{21} := \sum_{r=a+1}^{b} E\left[\frac{I(C\geq r)}{P(C\geq r|T_{r-1})}\frac{P(a\leq C\leq r-1|T_{r-1})}{P(a\leq C\leq b)}\left(E[m|T_r]-E[m|T_{r-1}]\right)D'\right],$$

$$B_{22} := \sum_{r=a+1}^{b} E\left[\frac{I(C\geq r)}{P(C\geq r|T_{r-1})}\frac{P(a\leq C\leq r-1|T_{r-1})}{P(a\leq C\leq b)}\left(E[m|T_r]-E[m|T_{r-1}]\right)\right.$$
$$\left.\times\sum_{k=1}^{R} I(C=k)\frac{\dot{P}(C=k|T_k)'}{P(C=k|T_k)}\right],$$

$$B_{31} := \sum_{r=a}^{b} E\left[\frac{I(C=r)}{P(a\leq C\leq b)}E[m|T_r]D'\right],$$

$$B_{32} := \sum_{r=a}^{b} E\left[\frac{I(C=r)}{P(a\leq C\leq b)}E[m|T_r]\sum_{k=1}^{R} I(C=k)\frac{\dot{P}(C=k|T_k)'}{P(C=k|T_k)}\right],$$

$$D := s(Z_1)+\sum_{k=2}^{R} I(C\geq k)s(Z_k|T_{k-1}).$$

We wrote this with the understanding that if $b=R$ then $B_{11}=B_{12}=0$, and if $a=b$ then $B_{21}=B_{22}=0$. For notational brevity define T_0 as any constant, so that $s(Z_1)\equiv s(Z_1|T_0)$. First, note that:

$$\begin{aligned}
B_{11} \;=\;& \sum_{r=b+1}^{R}\sum_{k=1}^{r} E\left[\frac{I(C\geq r)}{P(C\geq r|T_{r-1})}\frac{P(a\leq C\leq b|T_b)}{P(a\leq C\leq b)}\left(E[m|T_r]-E[m|T_{r-1}]\right)s(Z_k|T_{k-1})'\right]\\
&+\sum_{r=b+1}^{R}\sum_{k=r+1}^{R} E\left[\frac{I(C\geq k)}{P(C\geq r|T_{r-1})}\frac{P(a\leq C\leq b|T_b)}{P(a\leq C\leq b)}\left(E[m|T_r]-E[m|T_{r-1}]\right)s(Z_k|T_{k-1})'\right]\\
=\;& \sum_{r=b+1}^{R}\sum_{k=1}^{r} E\left[\frac{P(C\geq r|T_{r-1})}{P(C\geq r|T_{r-1})}\frac{P(a\leq C\leq b|T_b)}{P(a\leq C\leq b)}\left(E[m|T_r]-E[m|T_{r-1}]\right)s(Z_k|T_{k-1})'\right]\\
&+\sum_{r=b+1}^{R}\sum_{k=r+1}^{R} E\left[\frac{P(C\geq k|T_{k-1})}{P(C\geq r|T_{r-1})}\frac{P(a\leq C\leq b|T_b)}{P(a\leq C\leq b)}\left(E[m|T_r]-E[m|T_{r-1}]\right)s(Z_k|T_{k-1})'\right]
\end{aligned}$$

where the third and fourth lines follow by Lemma 9. Hence, we subsequently obtain that:

$$\begin{aligned} B_{11} &= \sum_{r=b+1}^{R} E\left[\frac{P(a \leq C \leq b|T_b)}{P(a \leq C \leq b)} E[m|T_r] s(Z_r|T_{r-1})'\right] + 0 \\ &= E\left[\frac{I(a \leq C \leq b)}{P(a \leq C \leq b)} m s(Z_R, \ldots, Z_{b+1}|T_b)'\right] \\ &= E\left[m s(Z_R, \ldots, Z_{b+1}|T_b)'|a \leq C \leq b\right]. \end{aligned} \tag{26}$$

The first equality follows since for all $k = 1, \ldots, r-1$: $E\left[(E[m|T_r] - E[m|T_{r-1}]) s(Z_k|T_{k-1})'\right] = E\left[E[(E[m|T_r] - E[m|T_{r-1}]) s(Z_k|T_{k-1})'|T_{r-1}]\right] = 0$ while for $k \geq r+1$: $E\left[E[m|T_r] s(Z_k|T_{k-1})'\right] = E\left[E[m|T_r] E[s(Z_k|T_{k-1})'|T_{k-1}]\right] = 0$. The second equality follows by (1) and Lemma 8 and the definition of score. The last equality is obvious.

Second, following the steps that led to the first line on the RHS of (26), we obtain that:

$$B_{21} = \sum_{r=a+1}^{b} E\left[\frac{P(a \leq C \leq r-1|T_{r-1})}{P(a \leq C \leq b)} E[m|T_{r-1}] s(Z_r|T_{r-1})'\right].$$

Therefore,

$$\begin{aligned} B_{21} &= \sum_{r=a+1}^{b} \sum_{k=a}^{r-1} E\left[\frac{P(C = k|T_k)}{P(a \leq C \leq b)} m s(Z_r|T_{r-1})'\right] \\ &= \sum_{r=a+1}^{b} \sum_{k=a}^{r-1} E\left[m s(Z_r|T_{r-1})'|C = k\right] \frac{P(C = k)}{P(a \leq C \leq b)} \\ &= \sum_{k=a}^{b-1} E\left[m \sum_{r=k+1}^{b} s(Z_r|T_{r-1})' \middle| C = k\right] \frac{P(C = k)}{P(a \leq C \leq b)} \\ &= \sum_{k=a}^{b-1} E\left[m s(Z_b, \ldots, Z_{k+1}|T_k)'|C = k\right] \frac{P(C = k)}{P(a \leq C \leq b)}. \end{aligned} \tag{27}$$

The first equality follows by (1) and Lemma 8. The second equality follows by the same steps that gave the second line on the RHS of (26). The third equality follows by interchanging the order of summations (allowed here). The last equality follows by the definition of score.

Third, we consider B_{31} and note that using the definition of score in the second equality below and the same argument as before in the third (last) equality below give:

$$\begin{aligned} B_{31} &= \sum_{r=a}^{b}\sum_{k=1}^{r} E\left[\frac{I(C=r)}{P(a\leq C\leq b)}E[m|T_r]s(Z_k|T_{k-1})'\right] \\ &= \sum_{r=a}^{b} E\left[\frac{I(C=r)}{P(a\leq C\leq b)}E[m|T_r]s(T_r)'\right] \\ &= \sum_{r=a}^{b} E\left[ms(T_r)'|C=r\right]\frac{P(C=r)}{P(a\leq C\leq b)}. \end{aligned} \tag{28}$$

Adding (27) and (28) gives:

$$\begin{aligned} B_{21}+B_{31} &= E\left[ms(T_b)'|C=b\right]\frac{P(C=b)}{P(a\leq C\leq b)} + \sum_{k=a}^{b-1} E\left[ms(T_b)'|C=k\right]\frac{P(C=k)}{P(a\leq C\leq b)} \\ &= E\left[ms(T_b)'|a\leq C\leq b\right]. \end{aligned} \tag{29}$$

Now, we consider the terms B_{12}, B_{22} and B_{32} respectively. Accordingly, first note that:

$$\begin{aligned} B_{12} &= \sum_{r=b+1}^{R}\sum_{k=r}^{R} E\left[\frac{I(C=k)}{P(C\geq r|T_{r-1})}\frac{P(a\leq C\leq b|T_b)}{P(a\leq C\leq b)}\left(E[m|T_r]-E[m|T_{r-1}]\right)\frac{\dot{P}(C=k|T_k)'}{P(C=k|T_k)}\right] \\ &= \sum_{r=b+1}^{R} E\left[\frac{1}{P(C\geq r|T_{r-1})}\frac{P(a\leq C\leq b|T_b)}{P(a\leq C\leq b)}\left(E[m|T_r]-E[m|T_{r-1}]\right)\sum_{k=r}^{R}\dot{P}(C=k|T_k)'\right] \\ &= \sum_{r=b+1}^{R} E\left[\frac{P(a\leq C\leq b|T_b)}{P(a\leq C\leq b)}\left(E[m|T_r]-E[m|T_{r-1}]\right)\frac{\dot{P}(C\geq r|T_{r-1})'}{P(C\geq r|T_{r-1})}\right] \\ &= 0. \end{aligned} \tag{30}$$

The second equality follows by (1) and Lemma 8. The third equality follows by Lemma 8 and Lemma 9. The fourth (last) equality follows by taking expectation conditional on T_{r-1} for the r-th term inside the summation. Exactly following the same steps as in the above (recall the analogy with B_{11} and B_{12} above) we obtain:

$$B_{22} = 0. \tag{31}$$

Lastly, as before, note that:

$$
\begin{aligned}
B_{32} &= \sum_{r=a}^{b} E\left[\frac{I(C=r)}{P(C=r|T_r)}\frac{E[m|T_r]\dot{P}(C=r|T_r)'}{P(a\leq C\leq b)}\right] = E\left[m\sum_{r=a}^{b}\frac{\dot{P}(C=r|T_r)'}{P(a\leq C\leq b)}\right] \\
&= E\left[\frac{P(a\leq C\leq b|T_b)}{P(a\leq C\leq b|T_b)}m\frac{\dot{P}(a\leq C\leq b|T_b)'}{P(a\leq C\leq b)}\right] = E\left[\frac{I(a\leq C\leq b)}{P(a\leq C\leq b|T_b)}m\frac{\dot{P}(a\leq C\leq b|T_b)'}{P(a\leq C\leq b)}\right] \\
&= E\left[m\frac{\dot{P}(a\leq C\leq b|T_b)'}{P(a\leq C\leq b|T_b)}\middle| a\leq C\leq b\right] \qquad (32)
\end{aligned}
$$

Therefore, (26) and (29)-(32) give the result. ■

A.2 Proof of the results stated in the main text

Proof of Lemma 1: For simplicity we suppress the dependence of quantities on O, Z, T_r, β, etc. unless confusing. Taking $a = 1, b = R$ in (3), note by using (4) and (5) that $\varphi_{[1,R]}(.)$ is:

$$
\begin{aligned}
& \sum_{j=1}^{R} P(C=j)\left\{\sum_{r=j+1}^{R}\frac{I(C\geq r)P(C=j|T_j)}{P(C=j)P(C\geq r|T_{r-1})}(E[m|T_r]-E[m|T_{r-1}]) + \frac{I(C=j)}{P(C=j)}E[m|T_j]\right\} \\
= & \sum_{r=2}^{R}\sum_{j=1}^{r-1} I(C\geq r)\frac{P(C=j|T_j)}{P(C\geq r|T_{r-1})}(E[m|T_r]-E[m|T_{r-1}]) + \sum_{j=1}^{R} I(C=j)E[m|T_j] \\
= & \sum_{r=2}^{R} I(C\geq r)\frac{P(C\leq r-1|T_{r-1})}{P(C\geq r|T_{r-1})}(E[m|T_r]-E[m|T_{r-1}]) + \sum_{r=1}^{R} I(C=r)E[m|T_r] \\
= & \sum_{r=2}^{R} I(C\geq r)\frac{1-P(C\geq r|T_{r-1})}{P(C\geq r|T_{r-1})}(E[m|T_r]-E[m|T_{r-1}]) + \sum_{r=1}^{R} I(C=r)E[m|T_r] \\
= & \sum_{r=2}^{R}\frac{I(C\geq r)}{P(C\geq r|T_{r-1})}(E[m|T_r]-E[m|T_{r-1}]) - \sum_{r=2}^{R} I(C\geq r)(E[m|T_r]-E[m|T_{r-1}]) + \sum_{r=1}^{R} I(C=r)E[m|T_r] \\
= & \sum_{r=2}^{R}\frac{I(C\geq r)}{P(C\geq r|T_{r-1})}(E[m|T_r]-E[m|T_{r-1}]) + E[m|T_1]
\end{aligned}
$$

where the last line follows because we can write $\sum_{r=2}^{R} I(C\geq r)(E[m|T_r]-E[m|T_{r-1}])$ as:

$$
I(C=R)E[m|T_R] + \sum_{r=2}^{R-1} E[m|T_r][I(C\geq r) - I(C\geq r+1)] + I(C\geq 2)E[m|T_1].
$$

The law of iterated expectations gives $E[\varphi_{[1,R]}(O;\beta)] = 0 + E[E[m(Z;\beta)|T_1]]$. Therefore,

$$\begin{aligned}
E[\varphi_{[1,R]}] &= E\left[\varphi_{[1,R]}\left\{\varphi_{[1,R]} - E[E[m(Z;\beta)|T_1]]\right\}'\right] \\
&= E\left[\sum_{r=2}^{R}\frac{I(C\geq r)}{P(C\geq r|T_{r-1})}(E[m|T_r]-E[m|T_{r-1}])\sum_{s=2}^{R}\frac{I(C\geq s)}{P(C\geq s|T_{s-1})}(E[m|T_s]-E[m|T_{s-1}])'\right] \\
&\quad +E[E[m|T_1](E[m|T_1]-E[E[m|T_1]])'] \\
&= \sum_{r=2}^{R}E\left[\frac{1}{P(C\geq r|T_{r-1})}E\left[(E[m|T_r]-E[m|T_{r-1}])(E[m|T_r]-E[m|T_{r-1}])'|T_{r-1}\right]\right] \\
&\quad +E[E[m|T_1](E[m|T_1]-E[m])'] \\
&= \sum_{r=2}^{R}E\left[\frac{V(E[m|T_r]|T_{r-1})}{P(C\geq r|T_{r-1})}\right] + V(E[m|T_1])
\end{aligned}$$

giving the desired result, where the last equality follows simply by definition while the third equality follows since for $r > s$ by using (1) and the law of iterated expectations:

$$\begin{aligned}
& E\left[\frac{I(C\geq r)}{P(C\geq r|T_{r-1})}(E[m|T_r]-E[m|T_{r-1}])\frac{I(C\geq s)}{P(C\geq s|T_{s-1})}(E[m|T_s]-E[m|T_{s-1}])'\right] \\
=\ & E\left[\frac{1-I(C\leq r-1)}{(1-P(C\leq r-1|T_{r-1}))P(C\geq s|T_{s-1})}(E[m|T_r]-E[m|T_{r-1}])(E[m|T_s]-E[m|T_{s-1}])'\right] \\
=\ & 0
\end{aligned}$$

and for $r > 1$, again by using (1) and the law of iterated expectations,

$$E\left[\frac{I(C\geq r)}{P(C\geq r|T_{r-1})}(E[m|T_r]-E[m|T_{r-1}])(E[m|T_1]-E[E[m|T_1]])'\right] = 0. \ \blacksquare$$

Remark: Unless confusing, we will write $m(Z;\beta^0_{[a,b]})$ simply as m for brevity in the sequel.

Proof of Proposition 2: We obtained the tangent set $\mathcal{T}$ in Section 3.2. In the just identified case the tangent set is given by (7) in the case of any generic $[a,b]$ with $a,b \in \{1,\ldots,R\}$ and $a \leq b$, while in the over identified case the tangent set is given by (7) and the additional restriction (10) if $a = 1, b = R$, and by (7) and the additional restriction (11) if $a = b$.

The rest of the proof will proceed as follows. We will show that $\varphi_{[a,b]}(O,\beta^0_{[a,b]})$ satisfies the pathwise derivative condition for any generic $[a,b]$ with $a, b \in \{1, \dots, R\}$ and $a \leq b$ in the over identified case ($d_m \geq d_\beta$). Thus, this will obviously be satisfied in the just identified case ($d_m = d_\beta$). Then we will show that the concerned influence function obtained from $\varphi_{[a,b]}(O, \beta^0_{[a,b]})$ belongs in $\mathcal{T}$ in the over identified case if $a = 1, b = R$ or if $a = b$, and belongs in $\mathcal{T}$ in the just identified case for any generic $[a,b]$ with $a, b \in \{1, \dots, R\}$ and $a \leq b$.

Taking any A that is a full row rank $d_\beta \times d_m$ matrix such that $AM_{[a,b]}$ is nonsingular, we know from Section 3.2 that:

$$\frac{\partial \beta^0_{[a,b]}(\eta_0)}{\partial \eta'} = -\left(AM_{[a,b]}\right)^{-1} AE\left[m(Z; \beta^0_{[a,b]}) \left\{ s(Z) + \frac{\dot{P}(a \leq C \leq b|T_b)}{P(a \leq C \leq b|T_b)} \right\}' \middle| a \leq C \leq b \right].$$

Therefore, the pathwise derivative condition

$$\frac{\partial \beta^0_{[a,b]}(\eta_0)}{\partial \eta'} = \left(AM_{[a,b]}\right)^{-1} AE\left[\varphi(O, \beta^0_{[a,b]}) S(O)'\right],$$

where $S(O)$ is defined in Section 3.2, will hold if:

$$E\left[\varphi_{[a,b]}(O, \beta^0_{[a,b]}) S(O)'\right] = E\left[m \left\{ s(Z) + \frac{\dot{P}(a \leq C \leq b|T_b)}{P(a \leq C \leq b|T_b)} \right\}' \middle| a \leq C \leq b \right],$$

and this is true by Lemma 10. The calculations for this demonstration are tedious and hence they are presented separately under Lemma 10 stated immediately before the present proof.

The pathwise derivative condition holds in the general over identified case ($d_m \geq d_\beta$). Hence, it also holds in the just identified case ($d_m = d_\beta$). To avoid any confusion (at the cost of brevity), we will first complete the proof for case (ii), i.e., the just identified case. We will show that the influence function $-M^{-1}_{[a,b]} \varphi_{[a,b]}(O, \beta^0_{[a,b]})$ obtained from $\varphi_{[a,b]}(O, \beta^0_{[a,b]})$ belongs in $\mathcal{T}$. This follows simply by matching the first set of terms in $-M^{-1}_{[a,b]} \varphi_{[a,b]}(O; \beta^0_{[a,b]})$ (i.e., those that correspond to line one in (25)) to the terms corresponding to $\nu_{b+1}(Z_1, \dots, Z_{b+1}), \dots, \nu_R(Z_1, \dots, Z_R)$ in $\mathcal{T}$; the second set of terms (i.e., those that correspond to line two in (25)) to the terms corresponding to $\nu_a(Z_1, \dots, Z_a), \dots, \nu_b(Z_1, \dots, Z_b)$

in $\mathcal{T}$; and the third set of terms (i.e., those that correspond to line three in (25)) to the terms corresponding to $\omega_a(Z_1, \ldots, Z_a), \ldots, \omega_b(Z_1, \ldots, Z_b)$ in $\mathcal{T}$; while matching zeros with the remaining terms in $\mathcal{T}$. Hence, $-M_{[a,b]}^{-1}\varphi_{[a,b]}(O;\beta_{[a,b]}^0)$ is the efficient influence function. The expectation of the outer-product of $-M_{[a,b]}^{-1}\varphi_{[a,b]}(O;\beta_{[a,b]}^0)$ gives the inverse efficiency bound $M_{[a,b]}^{-1}E\left[\varphi_{[a,b]}(O;\beta_{[a,b]}^0)\varphi_{[a,b]}'(O;\beta_{[a,b]}^0)\right]M_{[a,b]}^{-1\prime} = M_{[a,b]}^{-1}V_{[a,b]}M_{[a,b]}^{-1\prime}$.

Now let us get back to the over identified case ($d_m \geq d_\beta$). As noted in Section 3.2, this is where our proof markedly differs from similar proofs in the over identified case because those proofs only do a matching exercise similar to the above without considering the additional restrictions on the tangent set that are imposed by over identification. Arriving at the optimal A, i.e., $M_{[a,b]}'V_{[a,b]}^{-1}$, after this exercise is the same as in Chen et al. (2008) and hence to avoid repetition it is omitted for brevity.

First, consider the case of $a = b$. The above matching also holds with the influence function $-(M_{[a,a]}'V_{[a,a]}^{-1}M_{[a,a]})^{-1}M_{[a,a]}'V_{[a,a]}^{-1}\varphi_{[a,a]}(O,\beta_{[a,a]}^0)$. Hence, we focus on verifying the additional restriction (11) due to over identification. If $a = b$ then (11) is:

$$0 = B_{[a,a]}E\left[m(Z;\beta_{[a,a]}^0)\left\{\sum_{r=1}^{R}\nu_r(Z_1,\ldots,Z_r) + \omega_a(Z_1,\ldots,Z_a)\right\}'\middle| C = a\right].$$

Therefore, guided exactly by the above matching exercise, $-(M_{[a,a]}'V_{[a,a]}^{-1}M_{[a,a]})^{-1}M_{[a,a]}'V_{[a,a]}^{-1}\varphi_{[a,a]}(O,\beta_{[a,a]}^0)$ will satisfy (11) and hence belong in $\mathcal{T}$ if we can show that:

$$\begin{aligned} 0 \quad = \quad & B_{[a,a]}E\left[m\left\{\sum_{r=1}^{R}\frac{P(C=a|T_a)\left(E[m|T_r]-E[m|T_{r-1}]\right)}{P(C\geq r|T_{r-1})P(C=a)}\right.\right. \\ & \left.\left.+\frac{E[m|T_a]}{P(C=a)}\right\}'V_{[a,a]}^{-1}M_{[a,a]}(M_{[a,a]}'V_{[a,a]}^{-1}M_{[a,a]})^{-1}\middle| C=a\right]. \end{aligned}$$

Now, recalling that $B_{[a,b]} := \left(I_{d_\beta} - M_{[a,b]}\left(AM_{[a,b]}\right)^{-1}A\right)$, it follows that the above equation will hold if:

$$E\left[m\left\{\sum_{r=1}^{R}\frac{P(C=a|T_a)\left(E[m|T_r]-E[m|T_{r-1}]\right)}{P(C\geq r|T_{r-1})P(C=a)} + \frac{E[m|T_a]}{P(C=a)}\right\}'\middle| C=a\right] = V_{[a,a]},$$

which is true since by definition $V_{[a,a]} = E\left[\left\{\varphi_{[a,a]}(O;\beta^0_{[a,a]})\right\}\left\{\varphi_{[a,a]}(O;\beta^0_{[a,a]})\right\}'\right]$, i.e.,

$$\begin{aligned}
V_{[a,a]} &= E\left[\left\{\sum_{r=a+1}^{R}\frac{I(C\geq r)P(C=a|T_a)}{P(C\geq r|T_{r-1})P(C=a)}\left(E[m|T_r]-E[m|T_{r-1}]\right)+\frac{I(C=a)}{P(C=a)}E[m|T_a]\right\}\{\}'\right]\\
&= E\left[\sum_{r=a+1}^{R}\frac{P^2(C=a|T_a)}{P(C\geq r|T_{r-1})P^2(C=a)}m\left(E[m|T_r]-E[m|T_{r-1}]\right)'+\frac{I(C=a)}{P^2(C=a)}mE[m|T_a]'\right]\\
&= E\left[\sum_{r=a+1}^{R}\frac{I(C=a)P(C=a|T_a)}{P(C\geq r|T_{r-1})P^2(C=a)}m\left(E[m|T_r]-E[m|T_{r-1}]\right)'+\frac{I(C=a)}{P^2(C=a)}mE[m|T_a]'\right]\\
&= E\left[m\left\{\sum_{r=1}^{R}\frac{P(C=a|T_a)\left(E[m|T_r]-E[m|T_{r-1}]\right)}{P(C\geq r|T_{r-1})P(C=a)}+\frac{E[m|T_a]}{P(C=a)}\right\}'\middle| C=a\right].
\end{aligned}$$

Now, consider the case of $a=1, b=R$. The matching exercise from the just identified case will not be appropriate here because we have not imposed enough restrictions on the $\omega_r(Z_1,\ldots,Z_r)$'s; see footnote 10. Instead, here we will be guided by the simplified expression of $\varphi_{[1,R]}(O;\beta^0_{[1,R]})$ in Lemma 1, i.e.,

$$\varphi_{[1,R]}(O;\beta^0_{[1,R]}) = \sum_{r=2}^{R}\frac{I(C\geq r)}{P(C\geq r|T_{r-1})}\left(E[m|T_r]-E[m|T_{r-1}]\right)+E[m|T_1],$$

and match the term $-\left(M'_{[1,R]}V^{-1}_{[1,R]}M_{[1,R]}\right)^{-1}M'_{[1,R]}V^{-1}_{[1,R]}\frac{(E[m|T_r]-E[m|T_{r-1}])}{P(C\geq r|T_{r-1})}$ of the influence function $-\left(M'_{[1,R]}V^{-1}_{[1,R]}M_{[1,R]}\right)^{-1}M'_{[1,R]}V^{-1}_{[1,R]}\varphi_{[1,R]}(O;\beta^0_{[1,R]})$ with the term $\nu_r(Z_1,\ldots,Z_r)$ of $\mathcal{T}$ for $r=2,\ldots,R-1$ and the term $-\left(M'_{[1,R]}V^{-1}_{[1,R]}M_{[1,R]}\right)^{-1}M'_{[1,R]}V^{-1}_{[1,R]}E[m|T_1]$ of the influence function $-\left(M'_{[1,R]}V^{-1}_{[1,R]}M_{[1,R]}\right)^{-1}M'_{[1,R]}V^{-1}_{[1,R]}\varphi_{[1,R]}(O;\beta^0_{[1,R]})$ with the term $\nu_1(Z_1)$ of $\mathcal{T}$. Guided exactly by this matching exercise, $-(M'_{[1,R]}V^{-1}_{[1,R]}M_{[1,R]})^{-1}M'_{[1,R]}V^{-1}_{[1,R]}\varphi_{[1,R]}(O,\beta^0_{[1,R]})$ will satisfy (10) and hence belong in $\mathcal{T}$ if we can show that:

$$0 = B_{[1,R]}E\left[m\left\{\sum_{r=1}^{R}\frac{(E[m|T_r]-E[m|T_{r-1}])}{P(C\geq r|T_{r-1})}+E[m|T_1]\right\}'V^{-1}_{[1,R]}M_{[1,R]}(M'_{[1,R]}V^{-1}_{[1,R]}M_{[1,R]})^{-1}\right].$$

Now, recalling that $B_{[1,R]} := \left(I_{d_\beta}-M_{[1,R]}\left(AM_{[1,R]}\right)^{-1}A\right)$, it follows that the above equation will hold if:

$$E\left[m\left\{\sum_{r=1}^{R}\frac{(E[m|T_r]-E[m|T_{r-1}])}{P(C\geq r|T_{r-1})}+E[m|T_a]\right\}'\right]=V_{[1,R]},$$

which it is easy to see is true by following the same steps (but more easily) as done for the case $a=b$. Therefore, we have now established that for both cases $a=b$ and $a=1, b=R$, the influence function $-\Omega_{[a,b]}^{-1}M'_{[a,b]}V_{[a,b]}^{-1}\varphi_{[a,b]}(O;\beta_{[a,b]}^{0})$ belong in the tangent set $\mathcal{T}$. Therefore, $-\Omega_{[a,b]}^{-1}M'_{[a,b]}V_{[a,b]}^{-1}\varphi_{[a,b]}(O;\beta_{[a,b]}^{0})$ is the efficient influence function and hence the expectation of its outer-product gives the inverse efficiency bound $\Omega_{[a,b]}^{-1}$. ■

Proof of Proposition 3: The pathwise derivative condition for this result was verified in Chaudhuri (2020) for the just identified case and applies equally well to the over identified case (similar to what we saw in the proof of Proposition 2). Therefore, we only focus on characterizing the additional restrictions on the tangent set imposed by over identification, and showing that the claimed influence function satisfies those restrictions and thus is the efficient influence function. We hope that it will be clear along the process that the method in Section 3.2 to obtain the additional restrictions, is general enough to obtain the additional restriction on a tangent set that are imposed by over identification in other models too.

Proceeding exactly as in Section 3.2 but, importantly, reflecting the fact that $P(C=r|Z_1,\ldots,Z_r)$ is known, write the log of the distribution of O in terms of $(C,Z')'$ for a regular parametric sub-model indexed by η as (η was θ and Z_r was $Z_{(r)}$ in Chaudhuri (2020)):

$$\log f_\eta(O)=\log f_\eta(Z_1)+\sum_{r=2}^{R}I(C\geq r)\log f_\eta(Z_r|Z_1,\ldots,Z_{r-1})+\sum_{r=1}^{R}I(C=r)\log P(C=r|Z_1,\ldots,Z_r)$$

and the score function with respect to η as:

$$S_\eta(O)=s_\eta(Z_1)+\sum_{r=2}^{R}I(C\geq r)s_\eta(Z_r|Z_1,\ldots,Z_{r-1})$$

where $s_\eta(Z_1) := \frac{\partial}{\partial\eta}\log f_\eta(Z_1)$, and $s_\eta(Z_r|Z_1,\ldots,Z_{r-1}) := \frac{\partial}{\partial\eta}\log f_\eta(Z_r|Z_1,\ \ldots,Z_{r-1})$ for

$r = 2, \ldots, R$. The tangent set is the mean square closure of all d_β dimensional linear combinations of $S_\eta(O)$ for all such smooth parametric sub-models, and it can be generically defined as:

$$\mathcal{T} := \nu_1(Z_1) + \sum_{r=2}^{R} I(C \geq r)\nu_r(Z_1, \ldots, Z_r), \tag{33}$$

where $\nu_1(Z_1) \in L_0^2(F(Z_1))$ and $\nu_r(Z_1, \ldots, Z_r) \in L_0^2(F(Z_r|Z_1, \ldots, Z_{r-1}))$ for $r = 2, \ldots, R$.

We will proceed as before, but maintain that $P(C = r|Z_1, \ldots, Z_r)$ is known, to obtain the additional restrictions on $\mathcal{T}$ due to over identification. The moment restrictions in (12) give the following identity in η for a given λ:

$$0 \equiv E_\eta[m(Z; \beta_\lambda^0)|C \in \lambda] \equiv E_\eta\left[\frac{P(C \in \lambda|Z)}{P(a \leq C \leq b)} m(Z; \beta_\lambda^0)\right].$$

Differentiate it with respect to η under the integral at $\eta = \eta^0$, and use (1) and (12) to get:

$$0 = M_\lambda \frac{\partial \beta_\lambda^0(\eta_0)}{\partial \eta'} + E\left[m(Z; \beta_\lambda^0)s(Z)'\middle| C \in \lambda\right] \tag{34}$$

where $s(Z) := s(Z_1) + \sum_{r=2}^{R} s(Z_r|T_{r-1})$ (with abuse, we briefly revert to the T_r notation for brevity). Now, we note that (12) also gives the following identity in η for given λ:

$$0 \equiv AE_\eta[m(Z; \beta_\lambda^0)|C \in \lambda] \equiv AE_\eta\left[\frac{P(C \in \lambda|Z)}{P(C \in \lambda)} m(Z; \beta_\lambda^0)\right]$$

for any A that is a full row rank $d_\beta \times d_m$ matrix such that AM_λ is nonsingular. Then, as before, solving for $\frac{\partial \beta_\lambda^0(\eta_0)}{\partial \eta'}$, we obtain that:

$$\frac{\partial \beta_\lambda^0(\eta_0)}{\partial \eta'} = -\left(AM_\lambda\right)^{-1} AE\left[m(Z; \beta_\lambda^0)s(Z)'\middle| C \in \lambda\right],$$

which when substituted for in (34) gives (noting that $s(Z) := s(Z_1) + \sum_{r=2}^{R} s(Z_r|T_{r-1})$):

$$0 = \left(I_{d_\beta} - M_\lambda \left(AM_\lambda\right)^{-1} A\right) E\left[m(Z; \beta_\lambda^0)\left\{s(Z_1) + \sum_{r=2}^{R} s(Z_r|T_{r-1})\right\}'\middle| C \in \lambda\right].$$

While this is trivially true under just identification, in the case of over identification it implies that the tangent set $\mathcal{T}$ in (33) must satisfy the additional restrictions that

$$0 = \left(I_{d_\beta} - M_\lambda \left(AM_\lambda\right)^{-1} A\right) E\left[m(Z;\beta^0_\lambda)\left\{\nu(Z_1) + \sum_{r=2}^{R}\nu(Z_1,\ldots,Z_r)\right\}'\middle| C\in\lambda\right]. \quad (35)$$

Matching terms of $-\bar{\Omega}_\lambda^{-1}M'_\lambda\bar{V}_\lambda^{-1}\bar{\varphi}_\lambda(O;\beta^0_{[a,b]})$ with that of $\mathcal{T}$ where the term involving $\bar{\varphi}_{1,\lambda}$ is matched to $\nu_1(Z_1)$ and the terms involving $\frac{1}{P(C\geq r|T_{r-1})}\left(\bar{\varphi}_{r,\lambda} - \bar{\varphi}_{r-1,\lambda}\right)$ are matched to $\nu_r(Z_1,\ldots,Z_r)$ for $r=2,\ldots,R$, we can say that $-\bar{\Omega}_\lambda^{-1}M'_\lambda\bar{V}_\lambda^{-1}\bar{\varphi}_\lambda(O;\beta^0_{[a,b]})\in\mathcal{T}$ if additionally

$$0 = \left(I_{d_\beta} - M_\lambda \left(AM_\lambda\right)^{-1} A\right) E\left[m(Z;\beta^0_\lambda)\left\{\bar{\varphi}_{1,\lambda} + \sum_{r=2}^{R}\frac{\left(\bar{\varphi}_{r,\lambda} - \bar{\varphi}_{r-1,\lambda}\right)}{P(C\geq r|T_{r-1})}\right\}'\bar{V}_\lambda^{-1}M_\lambda\bar{\Omega}_\lambda^{-1}\middle| C\in\lambda\right]$$

which is true since we can easily see that, by repeatedly using (1) and the law of iterated expectations, we can write $\bar{V}_\lambda := E\left[\bar{\varphi}_\lambda(O;\beta^0_{[a,b]})\bar{\varphi}_\lambda(O;\beta^0_{[a,b]})'\right]$ as:

$$\begin{aligned}
\bar{V}_\lambda &= \sum_{r=2}^{R} E\left[\frac{\left(\bar{\varphi}_{r,\lambda} - \bar{\varphi}_{r-1,\lambda}\right)\left(\bar{\varphi}_{r,\lambda} - \bar{\varphi}_{r-1,\lambda}\right)'}{P(C\geq r|T_{r-1})}\right] + E\left[\bar{\varphi}_{1,\lambda}\bar{\varphi}'_{1,\lambda}\right] \\
&= \sum_{r=2}^{R} E\left[\frac{\bar{\varphi}_{r,\lambda}\left(\bar{\varphi}_{r,\lambda} - \bar{\varphi}_{r-1,\lambda}\right)'}{P(C\geq r|T_{r-1})}\right] + E\left[\bar{\varphi}_{1,\lambda}\bar{\varphi}'_{1,\lambda}\right] \\
&= \sum_{r=2}^{R} E\left[E\left[\frac{P(C\in\lambda|T_r)}{P(C\in\lambda)}m\middle| T_r\right]\frac{\left(\bar{\varphi}_{r,\lambda} - \bar{\varphi}_{r-1,\lambda}\right)'}{P(C\geq r|T_{r-1})}\right] + E\left[E\left[\frac{P(C\in\lambda|T_1)}{P(C\in\lambda)}m\middle| T_1\right]\bar{\varphi}'_{1,\lambda}\right] \\
&= \sum_{r=2}^{R} E\left[\frac{P(C\in\lambda|T_r)}{P(C\in\lambda)}m\frac{\left(\bar{\varphi}_{r,\lambda} - \bar{\varphi}_{r-1,\lambda}\right)'}{P(C\geq r|T_{r-1})}\right] + E\left[\frac{P(C\in\lambda|T_1)}{P(C\in\lambda)}m\bar{\varphi}'_{1,\lambda}\right] \\
&= \sum_{r=2}^{R} E\left[\frac{I(C\in\lambda)}{P(C\in\lambda)}m\frac{\left(\bar{\varphi}_{r,\lambda} - \bar{\varphi}_{r-1,\lambda}\right)'}{P(C\geq r|T_{r-1})}\right] + E\left[\frac{I(C\in\lambda)}{P(C\in\lambda)}m\bar{\varphi}'_{1,\lambda}\right] \\
&= E\left[m(Z;\beta^0_\lambda)\left\{\bar{\varphi}_{1,\lambda} + \sum_{r=2}^{R}\frac{\left(\bar{\varphi}_{r,\lambda} - \bar{\varphi}_{r-1,\lambda}\right)}{P(C\geq r|T_{r-1})}\right\}'\middle| C\in\lambda\right]. \blacksquare
\end{aligned}$$

Proof of Proposition 4: The pathwise derivative condition for this result was verified in Chaudhuri (2020) for the just identified case and applies equally well to the over identified case (similar to that in the proof of Proposition 2). Therefore, we only focus on characterizing the additional restrictions on the tangent set imposed by over identification, and showing that the claimed influence function satisfies those restrictions and thus is the efficient influence function. Proceeding as in Section 3.2 but imposing (13), write the log of the distribution of O in terms of $(C, Z')'$ for a regular parametric sub-model indexed by η as:

$$\log f_\eta(O) = \log f_\eta(Z_1) + \sum_{r=2}^{R} I(C \geq r) \log f_\eta(Z_r|Z_1, \ldots, Z_{r-1}) + \sum_{r=1}^{R} I(C = r) \log P(C = r|Z_1)$$

and the score function with respect to η as:

$$S_\eta(O) = s_\eta(Z_1) + \sum_{r=2}^{R} I(C \geq r) s_\eta(Z_r|Z_1, \ldots, Z_{r-1}) + \sum_{r=1}^{R} I(C = r) \frac{\dot{P}_\eta(C = r|Z_1)}{P_\eta(C = r|Z_1)}$$

where $s_\eta(Z_1) := \frac{\partial}{\partial\eta} \log f_\eta(Z_1)$, $s_\eta(Z_r|Z_1, \ldots, Z_{r-1}) := \frac{\partial}{\partial\eta} \log f_\eta(Z_r|Z_1, \ \ldots, Z_{r-1})$ for $r = 2, \ldots, R$, and $\dot{P}_\eta(C = r|Z_1) := \frac{\partial}{\partial\eta} P_\eta(C = r|Z_1)$ for $r = 1, \ldots, R$. We will actually use an apparently cumbersome but ultimately more convenient representation of the score function by using the two equivalent factorizations of the joint distribution of $I(C \in \lambda)$ and Z_1:

$$\begin{aligned} & s_\eta(Z_1) + I(C \in \lambda) \frac{\dot{P}_\eta(C \in \lambda|Z_1)}{P_\eta(C \in \lambda|Z_1)} + I(C \notin \lambda) \frac{\dot{P}_\eta(C \notin \lambda|Z_1)}{P_\eta(C \notin \lambda|Z_1)} \\ = {} & I(C \in \lambda) \left[\frac{\dot{P}_\eta(C \in \lambda)}{P_\eta(C \in \lambda)} + s_\eta(Z_1|C \in \lambda) \right] + I(C \notin \lambda) \left[\frac{\dot{P}_\eta(C \notin \lambda)}{P_\eta(C \notin \lambda)} + s_\eta(Z_1|C \notin \lambda) \right] \end{aligned} \quad (36)$$

where $s_\eta(Z_1|C \in \lambda) := \frac{\partial}{\partial\eta} \log f_\eta(Z_1|C \in \lambda)$, $s_\eta(Z_1|C \notin \lambda) := \frac{\partial}{\partial\eta} \log f_\eta(Z_1|C \notin \lambda)$, $\dot{P}_\eta(C \in \lambda|Z_1) := \frac{\partial}{\partial\eta} P_\eta(C \in \lambda|Z_1) =: -\dot{P}_\eta(C \notin \lambda|Z_1)$ and $\dot{P}_\eta(C \in \lambda) := \frac{\partial}{\partial\eta} P_\eta(C \in \lambda) =: -\dot{P}_\eta(C \notin \lambda)$. Then substituting for $s_\eta(Z_1)$ in $S_\eta(O)$ we obtain the cumbersome but useful expression:

$$\begin{aligned} S_\eta(O) \quad = \quad & I(C \in \lambda) \left[\frac{\dot{P}_\eta(C \in \lambda)}{P_\eta(C \in \lambda)} + s_\eta(Z_1|C \in \lambda) - \frac{\dot{P}_\eta(C \in \lambda|Z_1)}{P_\eta(C \in \lambda|Z_1)} \right] \\ & + I(C \notin \lambda) \left[\frac{\dot{P}_\eta(C \in \lambda)}{P_\eta(C \in \lambda) - 1} + s_\eta(Z_1|C \notin \lambda) - \frac{\dot{P}_\eta(C \in \lambda|Z_1)}{P_\eta(C \in \lambda|Z_1) - 1} \right] \\ & + \sum_{r=2}^{R} I(C \geq r) s_\eta(Z_r|Z_1, \ldots, Z_{r-1}) + \sum_{r=1}^{R} I(C = r) \frac{\dot{P}_\eta(C = r|Z_1)}{P_\eta(C = r|Z_1)}. \end{aligned}$$

Hence the representation of the tangent set that we will consider is:

$$\begin{aligned}\mathcal{T} &:= I(C\in\lambda)\left[\frac{a}{b}+\mu_1(Z_1,C\in\lambda)-\frac{a(Z_1)}{b(Z_1)}\right]+I(C\notin\lambda)\left[\frac{a}{b-1}+\mu_2(Z_1,C\notin\lambda)-\frac{a(Z_1)}{b(Z_1)-1}\right]\\ &\quad+\sum_{r=2}^{R}I(C\geq r)\nu_r(Z_1,\ldots,Z_r)+\sum_{r=1}^{R}I(C=r)\omega_r(Z_1), \qquad (37)\end{aligned}$$

where a and $b\in(0,1)$ are constants; $a(z_1)$ and $b(Z_1)$ are such that $a(Z_1)/b(Z_1)$ and $a(Z_1)/(b(Z_1)-1)$ are square integrable functions of Z_1; $\mu_1(Z_1,C\in\lambda)\in L_0^2(F(Z_1|C\in\lambda))$ and $\mu_2(Z_1,C\notin\lambda)\in L_0^2(F(Z_1|C\notin\lambda))$; and the terms described so far satisfy the restriction that the first line on the RHS of (37) is $L_0^2(F(Z_1))$ (since it represents $s(Z_1)$); whereas $\nu_r(Z_1,\ldots,Z_r)\in L_0^2(F(Z_r|Z_1,\ldots,Z_{r-1}))$ for $r=2,\ldots,R$, and $\omega_r(Z_1)$ is any square integrable function of Z_1 for $r=1,\ldots,R$. To obtain the additional restrictions due to over identification of β_λ^0, we write $\left(I_{d_\beta}-M_\lambda\left(AM_\lambda\right)^{-1}A\right)$ as B_λ for brevity, and then imposing CMAR in (13) we arrive at the counterpart of (9) for a given λ as:

$$0 = B_\lambda E\left[m(Z;\beta_\lambda^0)\left\{s(Z_1)+\sum_{r=2}^{R}s(Z_r|T_{r-1})+\frac{\dot{P}(C\in\lambda|Z_1)}{P(C\in\lambda|Z_1)}\right\}'\middle| C\in\lambda\right]$$

which gives the additional restrictions on $\mathcal{T}$ in (37) as:

$$0=B_\lambda\left\{E\left[m(Z;\beta_\lambda^0)\sum_{r=2}^{R}\nu_r(Z_1,\ldots,Z_r)'\middle|C\in\lambda\right]+E\left[m(Z;\beta_\lambda^0)\frac{I(C\in\lambda)}{P(C\in\lambda)}\left\{s(Z_1)+\frac{\dot{P}(C\in\lambda|Z_1)}{P(C\in\lambda|Z_1)}\right\}'\right]\right\}.$$

Substitute for $I(C\in\lambda)\left\{s(Z_1)+\dot{P}(C\in\lambda|Z_1)/P(C\in\lambda|Z_1)\right\}$ from (36) to get:

$$\begin{aligned}0 &= B_\lambda E\left[m\sum_{r=2}^{R}\nu_r(Z_1,\ldots,Z_r)'\middle|C\in\lambda\right]+B_\lambda E\left[m\frac{I(C\in\lambda)}{P(C\in\lambda)}\left\{s(Z_1|C\in\lambda)+\frac{\dot{P}(C\in\lambda)}{P(C\in\lambda)}\right\}'\right]\\ &= B_\lambda E\left[m\sum_{r=2}^{R}\nu_r(Z_1,\ldots,Z_r)'\middle|C\in\lambda\right]+B_\lambda E\left[m\frac{I(C\in\lambda)}{P(C\in\lambda)}s(Z_1|C\in\lambda)'\right] \qquad (38)\end{aligned}$$

using the moment restrictions in (12). (We are writing $m(Z;\beta_\lambda^0)$ as m for brevity.) Hence, over identification of β^0_λ imposes the additional restrictions (38) on $\mathcal{T}$ in (37).

Now, match the terms of $-\left[\Omega^{CMAR}_\lambda\right]^{-1} M'_\lambda \left[V^{CMAR}_\lambda\right]^{-1} \varphi^{CMAR}_\lambda(O;\beta^0_{[a,b]})$ with the terms of $\mathcal{T}$ as follows. The terms involving $\frac{P(C\in\lambda|T_1)}{P(C\geq r|T_{r-1})P(C\in\lambda)}\left(E[m|T_r]-E[m|T_{r-1}]\right)$ are matched to $\nu_r(Z_1,\ldots,Z_r)$ for $r=2,\ldots,R$. The term involving $\frac{I(C\in\lambda)}{P(C\in\lambda)}E[m|T_1]$ is matched to $I(C\in\lambda)s(Z_{(1)}|C\in\lambda)$. The other terms in $\mathcal{T}$ are matched to zeros. Therefore, the influence function $\left[\Omega^{CMAR}_\lambda\right]^{-1} M'_\lambda \left[V^{CMAR}_\lambda\right]^{-1} \varphi^{CMAR}_\lambda(O;\beta^0_{[a,b]})$ will belong in $\mathcal{T}$ and hence will be the efficient influence function if additionally:

$$\begin{aligned} 0 \;=\; & B_\lambda \left\{ E\left[m \sum_{r=2}^{R} \frac{P(C\in\lambda|T_1)}{P(C\geq r|T_{r-1})P(C\in\lambda)} \left(E[m|T_r]-E[m|T_{r-1}]\right)' \right| C\in\lambda \right] \\ & \qquad + E\left[m \frac{I(C\in\lambda)}{P^2(C\in\lambda)} E[m|T_1]' \right] \Bigg\} \left[V^{CMAR}_\lambda\right]^{-1} M_\lambda \left[\Omega^{CMAR}_\lambda\right]^{-1} \end{aligned}$$

i.e., if:

$$V^{CMAR}_\lambda = E\left[m \sum_{r=2}^{R} \frac{P^2(C\in\lambda|T_1)}{P(C\geq r|T_{r-1})P^2(C\in\lambda)} \left(E[m|T_r]-E[m|T_{r-1}]\right)' \right] + E\left[m \frac{I(C\in\lambda)}{P^2(C\in\lambda)} E[m|T_1]' \right],$$

which it can be seen is true by writing out the expression for $V^{CMAR}_\lambda := Var(\varphi^{CMAR}_\lambda(O;\beta^0_\lambda))$ and then using CMAR in (13) and the law of iterated expectations as in the last proof. ■

Proof of Lemma 5: Note that:

$$\begin{aligned} \omega^{\mathrm{IPW}}_{[a,b]} &:= \frac{I(C=R)}{\prod_{r=1}^{R-1}(1-P(C=r|T_r,C\geq r))} \frac{\sum_{j=a}^{b} P(C=j|T_j,C\geq j)\prod_{k=1}^{j-1}(1-P(C=k|T_k,C\geq k))}{P(a\leq C\leq b)} \\ &= \sum_{j=a}^{b} \frac{I(C=R)}{\prod_{r=1}^{R-1}(1-P(C=r|T_r,C\geq r))} \frac{P(C=j|T_j,C\geq j)\prod_{k=1}^{j-1}(1-P(C=k|T_k,C\geq k))}{P(a\leq C\leq b)} \\ &= \sum_{j=a}^{b} \frac{I(C=R)}{P(C=R|T_R)} \frac{P(C=j|T_j)}{P(a\leq C\leq b)} \qquad \left[= \sum_{j=a}^{b} \frac{P(C=j)}{P(a\leq C\leq b)} \frac{I(C=R)}{P(C=R|T_R)} \frac{P(C=j|T_j)}{P(C=j)} \right] \\ &= \sum_{j=a}^{b} \frac{I(C=R)}{P(C=R|T_R)} \frac{P(a\leq C\leq b|T_b)}{P(a\leq C\leq b)} = \sum_{j=a}^{b} \frac{I(C=R)}{P(C=R|T_R)} \frac{P(a\leq C\leq b|T_R)}{P(a\leq C\leq b)} \end{aligned}$$

where the last two equalities follow by (1). Therefore, since $Z\equiv T_R$, it follows by using the law of iterated expectations in the second and third equalities below, that:

$$\begin{aligned} E\left[\omega_{[a,b]}^{\text{IPW}} m(Z;\beta)\right] &= E\left[\frac{I(C=R)}{P(C=R|T_R)}\frac{P(a\le C\le b|T_R)}{P(a\le C\le b)}m(T_R;\beta)\right] \\ &= E\left[\frac{P(a\le C\le b|T_R)}{P(a\le C\le b)}m(T_R;\beta)\right] \\ &= E\left[\frac{I(a\le C\le b)}{P(a\le C\le b)}m(T_R;\beta)\right] \\ &= E[m(Z;\beta)|a\le C\le b]. \ \blacksquare \end{aligned}$$

Proof of Proposition 6: (i) will follow if Condition 1 of Ackerberg et al. (2014) holds. Our assumptions A1 and A3 directly imply Condition 1(i) and 1(ii) hold. Furthermore, Condition 1(iii) also holds by virtue of our assumption A2 because for any $r=a,\dots,R-1$:

$$\frac{\partial}{\partial p_r(T_r)}E\left[I(C\ge r)\left\{I(C=r)-p_r(T_r)\right\}|T_r\right] = -P(C\ge r|T_r)\ne 0 \quad \text{a.s. } T_r.$$

Before proceeding further, we note using the expression in (5) and Lemma 5 that:

$$I(C=R)\omega_{[a,b]}^{IPW}m(Z;\beta) = \sum_{j=a}^{b}\frac{P(C=j)}{P(a\le C\le b)}I(C=R)\omega_{[j,j]}^{IPW}m(Z;\beta); \tag{39}$$

and hence for the sake of a cleaner proof it is useful to work on:

$$I(C=R)\omega_{[j,j]}^{IPW}m(Z;\beta) \quad \text{where} \quad \omega_{[j,j]}^{IPW} = \frac{P(C=j|T_j,C\ge j)}{P(C=j)\prod_{k=j}^{R-1}[1-P(C=k|T_k,C\ge k)]}$$

and then combine the results based on the weights $P(C=j)/P(a\le C\le b)$.

For any $j=a,\dots,b$ replace $P(C=r|T_r,C\ge r)$ by $h_{j,r}(T_r):=1/(1-p_r(T_r))$ for $r=j+1,\dots,R-1$ and $P(C=j|T_j,C\ge j)$ by $h_{j,j}(T_j):=p_j(T_j)/(1-p_j(T_j))$ in $\omega_{[j,j]}^{IPW}$ to define (the reason behind the double subscript j,r in h will be clear soon):

$$\phi_{[j,j]}(C,T_R;\beta,h_{j,j}(T_j),\dots,h_{j,R-1}(T_{R-1})) := I(C=R)\frac{\prod_{k=j}^{R-1}h_{j,k}(T_j)}{P(C=j)}m(T_R;\beta). \tag{40}$$

Let $h^0_{j,r}(T_r) := 1/(1 - P(C = r|T_r, C \geq r))$ for $r = j+1, \ldots, R-1$ and $h^0_{j,j}(T_j) = P(C = j|T_j, C \geq j)/(1 - P(C = j|T_j, C \geq j))$. Then, trivially $\frac{\partial E[\phi_{[j,j]}(C,T_R;\beta,h^0_{j,j}(T_j),\ldots,h^0_{j,R-1}(T_{R-1}))]}{\partial h_{j,r}}[.]$ is a linear functional for $r = j, \ldots, R-1$. We maintain the assumption that it is also a bounded functional as defined in Ackerberg et al. (2014). (The boundedness is maintained as a high level assumption since under our assumption A2 it can hold in various ways depending on the interplay between the $E[m|T_r]$'s and the conditional hazards; e.g., taking $j = 1, R = 2$, we can see that $\frac{\partial E[\phi_{[1,1]}(C,T_2;\beta,h_{1,1})]}{\partial h_{1,1}} = E[P(C = 2|T_1)m(Z; \beta^0_{[1,R]})/P(C = 1)]$.) Thus, Condition 1(iv) of Ackerberg et al. (2014) also holds under our maintained assumptions. However, our interest is not always on a unitary sub-population $[j, j]$ but more generally on $[a, b]$, and for that we know from (39) that we should be looking at:

$$\sum_{j=a}^{b} \frac{P(C = j)}{P(a \leq C \leq b)} \phi_{[j,j]}(C, T_R; \beta, h_{j,j}(T_j), \ldots, h_{j,R-1}(T_{R-1})).$$

Before proceeding further we remark here about the double subscript in h. We redefined the nuisance parameters as h to make the functionals linear in h. However, the h's that enter the above linear combination are not unique — $h_{a,k}(T_k) = \ldots = h_{k-1,k}(T_k)$ for any $k = a+1, \ldots, b$ and $h_{a,k}(T_k) = \ldots = h_{k-1,k}(T_k)$ for any $k = b+1, \ldots, R-1$, while $h_{j,k}(T_k)$ appearing in $\phi_{[j,j]}(.)$ and $h_{k,k}(T_k)$ appearing in $\phi_{[k,k]}(.)$ for $k = j+1, \ldots, b$ and $j = a \ldots, b-1$ both depend on $p_k(T_k)$ only but in different ways. Pretending that the h's are distinct does not cause any problem, not even with the invertibility in Condition 1(iii) of Ackerberg et al. (2014) since that will lead to a diagonal matrix (and it will be important to keep this last statement in mind for the proof of part (ii)). Therefore if Condition 1 of Ackerberg et al. (2014) holds for $\phi_{[j,j]}(.)$ for $j = a, \ldots, b$, which we have already shown, then it also holds for the above linear combination those $\phi_{[j,j]}(.)$'s. This completes the proof of part (i).

(ii) It is straightforward to see that:

$$E\left[\frac{\partial}{\partial \beta'} \sum_{j=a}^{b} \frac{P(C = j)}{P(a \leq C \leq b)} \phi_{[j,j]}(C, T_R; \beta^0_{[a,b]}, h_{j,j}(T_j), \ldots, h_{j,R-1}(T_{R-1}))\right] = M_{[a,b]}.$$

Hence, we know from Theorem 1 of Ackerberg et al. (2014) that the efficiency bound for $\beta^0_{[a,b]}$ based on the information contained in *only* the moment restrictions in part (i) is:

$$\tilde{\Omega}_{[a,b]} := M'_{[a,b]} \left[Var \left(\sum_{j=a}^{b} \frac{P(C=j)}{P(a \leq C \leq b)} \widetilde{\phi}_{[j,j]}(C, T_R; \beta^0_{[a,b]}, h^0_{j,j}(T_j), \ldots, h^0_{j,R-1}(T_{R-1})) \right) \right]^{-1} M_{[a,b]} \tag{41}$$

where, writing $(h_{j,j}(T_j), \ldots, h_{j,R-1}(T_{R-1}))$ as $h_{j,j:R-1}(T_{R-1})$ and its true value as $h^0_{j,j:R-1}(T_{R-1})$:

$$\widetilde{\phi}_{[j,j]}(C, T_R; \beta, h_{j,j:R-1}(T_{R-1})) := \phi_{[j,j]}(C, T_R; \beta, h_{j,j:R-1}(T_{R-1})) - \sum_{k=j}^{R-1} \frac{D^0_{j,k}(T_k;\beta)}{S^0_{j,k}(T_k)} s_{j,k}(C, T_k, h_{j,k}(T_k)) \tag{42}$$

and where $D^0_{j,k}(T_k;\beta)$, $S^0_{j,k}(T_k)$ and $s_{j,k}(C, T_k, h_{j,k}(T_k))$ are as follows. For $j = a, \ldots, b$:

$$\begin{aligned} s_{j,k}(C, T_k, h_{j,k}(T_k)) :=&\ I(C \geq k) \left[I(C=k) - \frac{h_{j,k}(T_k) - 1}{h_{j,k}(T_k)} \right] \quad \text{for } k = j+1, \ldots, R-1 \\ :=&\ I(C \geq k) \left[I(C=k) - \frac{h_{j,k}(T_k)}{1 + h_{j,k}(T_k)} \right] \quad \text{for } k = j \end{aligned}$$

whereas for $j = a, \ldots, b$ and $k = j, \ldots, R-1$:

$$S^0_{j,k}(T_k) := \frac{\partial E[s_{j,k}(C, T_k, h^0_{j,k}(T_k))]}{\partial h_{j,k}} = -P(C \geq |k|T_k) \left(1 - P(C = k|T_k, C \geq k)\right)^2.$$

$D^0_{j,k}(T_k;\beta) v_{j,k}(T_k)$ is the pathwise derivative of $E[\phi_{[j,j]}(C, T_R; \beta, h_{j,j:R-1}(T_{R-1}))|T_k]$ with respect to $h_{j,k}(T_k)$ in the direction $v_{j,k}(T_k) \in H_{j,k}(T_k) - \{h^0_{j,k}(T_k)\}$ (where $\mathcal{H}_{j,k}(T_k)$ is the function space for $h_{j,k}(T_k)$) evaluated at $h^0_{j,j:R-1}(T_{R-1})$, i.e., for $j = a, \ldots, b$ and $k = j, \ldots, R-1$:

$$D^0_{j,k}(T_k;\beta) v_{j,k}(T_k) = \frac{\partial E[\phi_{[j,j]}(C, T_R; \beta, h^0_{j,j:R-1}(T_{R-1}))|T_k]}{\partial h_{j,k}} [v_{j,k}].$$

First, note that:

$$D^0_{j,k}(T_k;\beta) = E\left[\left\{ \prod_{r=j,\ldots,R-1; r \neq k} h_{j,r}(T_r) \right\} I(C=R) \frac{m(Z;\beta)}{P(C=j)} \middle| T_k \right]$$

i.e., for $k = j+1, \ldots, R-1$:

$$
\begin{aligned}
D^0_{j,k}(T_k;\beta) &= E\left[\left\{\prod_{r=j}^{R-1} h_{j,r}(T_r)\right\} I(C=R)\frac{m(Z;\beta)}{P(C=j)h_{j,k}(T_k)}\right| T_k\Bigg] \\
&= E\left[\frac{I(C=R)P(C=j|T_j, C\geq j)}{\prod_{r=j}^{R-1}(1-P(C=r|T_r, C\geq r))}\frac{m(Z;\beta)}{P(C=j)h_{j,k}(T_k)}\right| T_k\Bigg] \\
&= E\left[\frac{I(C=R)P(C=j|T_j, C\geq j)\prod_{r=1}^{j-1}(1-P(C=r|T_r, C\geq r))}{\prod_{r=1}^{R-1}(1-P(C=r|T_r, C\geq r))}\frac{m(Z;\beta)}{P(C=j)h_{j,k}(T_k)}\right| T_k\Bigg] \\
&= E\left[\frac{I(C=R)P(C=j|T_j)}{P(C=R|T_{R-1})}\frac{m(Z;\beta)}{P(C=j)h_{j,k}(T_k)}\right| T_k\Bigg] \\
&= E\left[\frac{m(Z;\beta)P(C=j|T_j)}{P(C=j)h_{j,k}(T_k)}\right| T_k\Bigg] \\
&= \frac{E[m(Z;\beta)|T_k]P(C=j|T_j)(1-P(C=k|T_k, C\geq k))}{P(C=j)}
\end{aligned}
$$

where the second last equality follows by (1) and the law of iterated expectations, whereas:

$$
\begin{aligned}
D^0_{j,j}(T_j;\beta) &= E\left[\left\{\prod_{r=j+1}^{R-1} h_{j,r}(T_r)\right\} I(C=R)\frac{m(Z;\beta)}{P(C=j)}\right| T_j\Bigg] \\
&= E\left[\frac{I(C=R)}{\prod_{r=j+1}^{R-1}(1-P(C=r|T_r, C\geq r))}\frac{m(Z;\beta)}{P(C=j)}\right| T_j\Bigg] \\
&= E\left[\frac{I(C=R)\prod_{r=1}^{j}(1-P(C=r|T_r, C\geq r))}{\prod_{r=1}^{R-1}(1-P(C=r|T_r, C\geq r))}\frac{m(Z;\beta)}{P(C=j)}\right| T_j\Bigg] \\
&= E\left[\frac{I(C=R)P(C\geq j+1|T_j)}{P(C=R|T_{R-1})}\frac{m(Z;\beta)}{P(C=j)}\right| T_j\Bigg] \\
&= E\left[P(C\geq j+1|T_j)\frac{m(Z;\beta)}{P(C=j)}\right| T_j\Bigg] \\
&= \frac{E[m(Z;\beta)|T_j]P(C\geq j+1|T_j)}{P(C=j)}
\end{aligned}
$$

where, as before, the second last equality follows by (1) and the law of iterated expectations.

Plugging them in (42) at $\beta^0_{[a,b]}, h^0_{j,j:R-1}(T_{R-1})$ gives:

$$
\widetilde{\phi}_{[j,j]}(C, T_R; \beta^0_{[a,b]}, h^0_{j,j:R-1}(T_{R-1})) = \phi_{[j,j]}(C, T_R; \beta^0_{[a,b]}, h^0_{j,j:R-1}(T_{R-1})) - \sum_{k=j}^{R-1}\frac{D^0_{j,k}(T_k;\beta)}{S^0_{j,k}(T_k)} s_{j,k}(C, T_k, h^0_{j,k}(T_k))
$$

where, writing $m(Z;\beta^0_{[a,b]})$ as m for brevity, the RHS of the above equation is:

$$
\begin{aligned}
& I(C=R)\omega^{IPW}_{[j,j]} m - \sum_{k=j+1}^{R-1} \frac{D^0_{j,k}(T_k;\beta)}{S^0_{j,k}(T_k)} s_{j,k}(C,T_k,h^0_{j,k}(T_k)) - \frac{D^0_{j,j}(T_k;\beta)}{S^0_{j,j}(T_k)} s_{j,j}(C,T_k,h^0_{j,j}(T_j)) \\
= \; & \varphi_{[j,j]}(O;\beta^0_{[a,b]}) \qquad (43)
\end{aligned}
$$

by (6) because we know from the above calculations that for $k = j+1, \ldots, R-1$:

$$
\begin{aligned}
& -\frac{D^0_{j,k}(T_k;\beta)}{S^0_{j,k}(T_k)} s_{j,k}(C,T_k,h^0_{j,k}(T_k)) \\
= \; & \frac{E[m|T_k]\frac{P(C=j|T_j)}{P(C=j)}(1-P(C=k|T_k,C\geq k))}{P(C\geq k|T_k)(1-P(C=k|T_k,C\geq k))^2}\left[I(C=k) - I(C\geq k)P(C=k|T_k,C\geq k)\right] \\
= \; & \frac{E[m|T_k]\frac{P(C=j|T_j)}{P(C=j)}}{P(C\geq k|T_k)(1-P(C=k|T_k,C\geq k))}\left[I(C\geq k) - I(C\geq k+1) - I(C\geq k)P(C=k|T_k,C\geq k)\right] \\
= \; & \left[\frac{I(C\geq k)}{P(C\geq k|T_k)} - \frac{I(C\geq k+1)}{P(C\geq k|T_k)(1-P(C=k|T_k,C\geq k))}\right]\frac{P(C=j|T_j)}{P(C=j)}E[m|T_j] \\
= \; & \left[\frac{I(C\geq k)}{P(C\geq k|T_k)} - \frac{I(C\geq k+1)}{P(C\geq k+1|T_k)}\right]\frac{P(C=j|T_j)}{P(C=j)}E[m|T_j] \\
= \; & \left[\frac{I(C\geq k)}{P(C\geq k|T_{k-1})} - \frac{I(C\geq k+1)}{P(C\geq k+1|T_k)}\right]\frac{P(C=j|T_j)}{P(C=j)}E[m|T_j] \qquad \text{[by Lemma 8]}
\end{aligned}
$$

whereas for $k = j$:

$$
\begin{aligned}
& -\frac{D^0_{j,j}(T_j;\beta)}{S^0_{j,j}(T_j)} s_{j,j}(C,T_j,h^0_{j,j}(T_j)) \\
= & \frac{E[m|T_j]\frac{P(C\geq j+1|T_j)}{P(C=j)}}{P(C\geq j|T_j)(1-P(C=j|T_j,C\geq j))^2}\left[I(C=j) - I(C\geq j)P(C=j|T_j,C\geq j)\right] \\
= & \frac{E[m|T_j]\frac{P(C\geq j+1|T_j)}{P(C=j)}}{P(C\geq j|T_j)(1-P(C=j|T_j,C\geq j))^2}\left[I(C=j) - \{I(C=j)+I(C\geq j+1)\}P(C=j|T_j,C\geq j)\right] \\
= & \left[\frac{I(C=j)P(C\geq j+1|T_j)}{P(C\geq j|T_j)(1-P(C=j|T_j,C\geq j))} - \frac{I(C\geq j+1)P(C\geq j+1|T_j)P(C=j|T_j,C\geq j)}{P(C\geq j|T_j)(1-P(C=j|T_j,C\geq j))^2}\right]\frac{E[m|T_j]}{P(C=j)} \\
= & \left[\frac{I(C=j)P(C\geq j+1|T_j)}{P(C\geq j+1|T_j)} - \frac{I(C\geq j+1)P(C\geq j+1|T_j)\frac{P(C=j|T_j)}{P(C\geq j|T_j)}}{P(C\geq j+1|T_j)\frac{P(C\geq j|T_j)-P(C=j|T_j)}{P(C\geq j|T_j)}}\right]\frac{E[m|T_j]}{P(C=j)} \\
= & \left[I(C=j) - I(C\geq j+1)\frac{P(C=j|T_j)}{P(C\geq j+1|T_j)}\right]\frac{E[m|T_j]}{P(C=j)}.
\end{aligned}
$$

Therefore, using (43) and (6) for the first equality and then (3) for the second, imply that:

$$\sum_{j=a}^{b} \frac{P(C=j)}{P(a \le C \le b)} \widetilde{\phi}_{[j,j]}(C, T_R; \beta^0_{[a,b]}, h^0_{j,j:R-1}(T_{R-1})) = \sum_{j=a}^{b} \frac{P(C=j)}{P(a \le C \le b)} \varphi_{[j,j]}(O; \beta^0_{[a,b]}) = \varphi_{[a,b]}(O; \beta^0_{[a,b]}).$$

Hence, by (41) we obtain that $\tilde{\Omega}_{[a,b]} = \Omega_{[a,b]}$ as defined in Proposition 2. ■

Proof of Lemma 7: (i) The equivalence of the limited and full information approach here follows exactly as in part (i) of Proposition 6, with the only change that the conditional hazards are all now conditioned on T_1 only. Consequently, the influence function in part (i) will take a different form here, and for the rest of the proof of part (i) we derive that form. To avoid introducing new notation we follow the notation from the last proposition as much as we can. We know from Theorem 1 of Ackerberg et al. (2014) that the efficiency bound for $\beta^0_{[a,b]}$ based on the information contained *only* in the moment restrictions in part (i) is:

$$M'_{[a,b]} \left[Var \left(\sum_{j=a}^{b} \frac{P(C=j)}{P(a \le C \le b)} \widetilde{\phi}_{[j,j]}(C, T_R; \beta^0_{[a,b]}, h^0_{j,j}(T_1), \ldots, h^0_{j,R-1}(T_1)) \right) \right]^{-1} M_{[a,b]}$$

where, writing $(h_{j,j}(T_1), \ldots, h_{j,R-1}(T_1))$ as $h_{j,j:R-1}(T_1)$ and its true value as $h^0_{j,j:R-1}(T_1)$:

$$\begin{aligned}
\widetilde{\phi}_{[j,j]}(C, T_R; \beta, h_{j,j:R-1}(T_1)) &:= \phi_{[j,j]}(C, T_R; \beta, h_{j,j:R-1}(T_1)) - \sum_{k=j}^{R-1} \frac{D^0_{j,k}(T_1; \beta)}{S^0_{j,k}(T_1)} s_{j,k}(C, T_1, h_{j,k}(T_1)) \\
\phi_{[j,j]}(C, T_R; \beta, h_{j,j:R-1}(T_1)) &:= I(C=R) \frac{\prod_{k=j}^{R-1} h_{j,k}(T_1)}{P(C=j)} m(T_R; \beta) \\
s_{j,k}(C, T_1, h_{j,k}(T_1)) &:= I(C \ge k) \left[I(C=k) - \frac{h_{j,k}(T_1) - 1}{h_{j,k}(T_1)} \right] \quad \text{for } k = j+1, \ldots, R-1 \\
&:= I(C \ge k) \left[I(C=k) - \frac{h_{j,k}(T_1)}{1 + h_{j,k}(T_1)} \right] \quad \text{for } k = j \\
S^0_{j,k}(T_1) &:= \frac{\partial E[s_{j,k}(C, T_1, h^0_{j,k}(T_1))]}{\partial h_{j,1}} = -P(C \ge |k|T_1) \left(1 - P(C=k|T_1, C \ge k)\right)^2
\end{aligned}$$

for $j = a, \ldots, b$ and $k = j, \ldots, R-1$. $D^0_{j,k}(T_1; \beta) v_{j,k}(T_1)$ is the pathwise derivative of $E[\phi_{[j,j]}(C, T_R; \beta, h_{j,j:R-1}(T_1))|T_1]$ with respect to $h_{j,k}(T_1)$ in the direction $v_{j,k}(T_1) \in H_{j,k}(T_1)-$

$\{h^0_{j,k}(T_1)\}$ (where $\mathcal{H}_{j,k}(T_1)$ is the function space for $h_{j,k}(T_1)$) evaluated at $h^0_{j,j:R-1}(T_1)$, i.e.,

$$D^0_{j,k}(T_1;\beta)v_{j,k}(T_1) = \frac{\partial E[\phi_{[j,j]}(C,T_R;\beta,h^0_{j,j:R-1}(T_1))|T_1]}{\partial h_{j,k}}[v_{j,k}] \quad \text{for } j=a,\ldots,b \text{ and } k=j,\ldots,R-1.$$

Therefore, just like before (but now with conditioning set T_1 for all terms):

$$\begin{aligned} D^0_{j,k}(T_1;\beta) &= \frac{E[m(Z;\beta)|T_1]P(C=j|T_1)(1-P(C=k|T_1,C\geq k))}{P(C=j)} \quad \text{for } k=j+1,\ldots,R-1, \\ D^0_{j,j}(T_1;\beta) &= \frac{E[m(Z;\beta)|T_1]P(C\geq j+1|T_1)}{P(C=j)}, \end{aligned}$$

and hence for $k=j+1,\ldots,R-1$:

$$-\frac{D^0_{j,k}(T_1;\beta)}{S^0_{j,k}(T_1)}s_{j,k}(C,T_1,h^0_{j,k}(T_1)) = \left[\frac{I(C\geq k)}{P(C\geq k|T_1)} - \frac{I(C\geq k+1)}{P(C\geq k+1|T_1)}\right]\frac{P(C=j|T_1)}{P(C=j)}E[m|T_1]$$

whereas for $k=j$:

$$-\frac{D^0_{j,j}(T_1;\beta)}{S^0_{j,j}(T_1)}s_{j,j}(C,T_1,h^0_{j,j}(T_1)) = \left[I(C=j) - I(C\geq j+1)\frac{P(C=j|T_1)}{P(C\geq j+1|T_1)}\right]\frac{E[m|T_1]}{P(C=j)},$$

and therefore:

$$-\sum_{k=j}^{R-1}\frac{D^0_{j,k}(T_1;\beta)}{S^0_{j,k}(T_1)}s_{j,k}(C,T_1,h^0_{j,k}(T_1)) = \left\{\frac{I(C=j)}{P(C=j)} - \frac{I(C=R)}{P(C=R|T_1)}\frac{P(C=j|T_1)}{P(C=j)}\right\}E[m|T_1],$$

which gives:

$$\begin{aligned} &\widetilde{\phi}_{[j,j]}(C,T_R;\beta^0_{[a,b]},h^0_{j,j:R-1}(T_1)) \\ &= \frac{I(C=R)}{P(C=R|T_1)}\frac{P(C=j|T_1)}{P(C=j)}m + \left\{\frac{I(C=j)}{P(C=j)} - \frac{I(C=R)}{P(C=R|T_1)}\frac{P(C=j|T_1)}{P(C=j)}\right\}E[m|T_1]. \end{aligned}$$

Therefore,

$$\sum_{j=a}^{n}\frac{P(C=j)}{P(a\leq C\leq b)}\widetilde{\phi}_{[j,j]}(C,T_R;\beta^0_{[a,b]},h^0_{j,j:R-1}(T_1)) = \varphi^\dagger_{[a,b]}.$$

Adding and subtracting the same terms to $\varphi^{\dagger}_{[a,b]}$ in order to match $\varphi^{CMAR}(O;\beta^0_{[a,b]})$ from Proposition 4, we obtain:

$$\begin{aligned}
\varphi^{\dagger}_{[a,b]} &= \sum_{r=2}^{R} \frac{I(C \geq R)}{P(C \geq R|T_1)} \frac{P(a \leq C \leq b|T_1)}{P(a \leq C \leq b)} \left(E[m|T_r] - E[m|T_{r-1}]\right) \\
&\quad + \left[\frac{I(a \leq C \leq b)}{P(a \leq C \leq b)} - \left(\frac{I(C \geq R)}{P(C \geq R|T_1)} - \frac{I(C \geq R)}{P(C \geq R|T_1)}\right) \frac{P(a \leq C \leq b|T_1)}{P(a \leq C \leq b)}\right] E[m|T_1] \\
&= \sum_{r=2}^{R} \frac{I(C \geq R)}{P(C \geq R|T_1)} \frac{P(a \leq C \leq b|T_1)}{P(a \leq C \leq b)} \left(E[m|T_r] - E[m|T_{r-1}]\right) + \frac{I(a \leq C \leq b)}{P(a \leq C \leq b)} E[m|T_1].
\end{aligned}$$

(ii) Taking variance, we obtain:

$$V^{\dagger}_{[a,b]} = \sum_{r=2}^{R} E\left[\frac{P^2(a \leq C \leq b|T_1)}{P(C \geq R|T_1) P^2(a \leq C \leq b)} Var\left(E[m|T_r]|T_{r-1}\right)\right] + E\left[\frac{I(a \leq C \leq b)}{P^2(a \leq C \leq b)} E[m|T_1] E'[m|T_1]\right]$$

whereas we know from Proposition 4 that:

$$V^{CMAR}_{[a,b]} = \sum_{r=2}^{R} E\left[\frac{P^2(a \leq C \leq b|T_1)}{P(C \geq r|T_1) P^2(a \leq C \leq b)} Var\left(E[m|T_r]|T_{r-1}\right)\right] + E\left[\frac{I(a \leq C \leq b)}{P^2(a \leq C \leq b)} E[m|T_1] E'[m|T_1]\right]$$

Therefore, we obtain that $V^{\dagger}_{[a,b]} - V^{CMAR}_{[a,b]}$ is:

$$\begin{aligned}
&\sum_{r=2}^{R} E\left[\frac{P^2(a \leq C \leq b|T_1)}{P^2(a \leq C \leq b)} \left[\frac{1}{P(C \geq R|T_1)} - \frac{1}{P(C \geq r|T_1)}\right] Var\left(E[m|T_r]|T_{r-1}\right)\right] \\
= &\sum_{r=2}^{R} E\left[\frac{P(a \leq C \leq b|T_1)}{P(a \leq C \leq b)} \left[\frac{1}{P(C \geq R|T_1)} - \frac{1}{P(C \geq r|T_1)}\right] Var\left(E[m|T_r]|T_{r-1}\right) \middle| a \leq C \leq b\right],
\end{aligned}$$

which is positive semi-definite by construction. ■

Remark: The results also hold if the moment restrictions in Lemma 7(i) are replaced by:

$$E\left[\frac{I(C=R)}{p_R(T_1)} \frac{p_{[a,b]}(T_1)}{P(a \leq C \leq b)} m(Z;\beta)\right] = 0 \quad \text{and} \quad E\left[\left(\begin{array}{c} I(C=R) - p_R(T_1) \\ I(a \leq C \leq b) - p_{[a,b]}(T_1) \end{array}\right) \middle| T_1\right] = 0$$

almost surely T_1. This representation is also usable in practice since T_1 is always observed.

B Supplemental Appendix B: Monte Carlo experiment

We will now study the small-sample properties of our proposed estimator EFF and inference based on it for estimands that are similar to those considered in our empirical illustration.

B.1 Simulation design

We will consider a setup reflecting the individual's decision to stay or leave dynamically over periods from programs (e.g., smoking cessation, weight loss), school, job, marriage, experiments, surveys, market, etc. We model this decision to leave after any period as a simple comparison between the individual's expectation of the outcome and their actual outcome after that period. Accordingly, we will consider an R-period program where Y_r is the outcome from staying until the end of the r-th period for $r = 1, \ldots, R$ in the program. We will assume that this outcome is generated as follows. For $t = 1, \ldots, T$, let:

$$Y_t = |Y_{t-1}| + Y_{t-2} + X_t + e_t, \quad \text{where} \quad X_t = X_{t-1} + v_t.$$

e_t and v_t are the model errors.[20] Take X_0, Y_{-1}, Y_0 independently $N(1,1)$ as the initial state. Our analysis below is not conditional on the initial state, but this could be done. We will take $R = T = 3$, and let X_r be the other observed variables for the r-th period for $r = 1, \ldots, R$.

Let the individual's expectation for the outcome in the r-th period be Y_r^*. Suppose that the individual decides to leave the program at the end of the r-th period, conditional on staying until then, if and only if the actual outcome exceeds the expectation, i.e., $Y_r^* < Y_r$. In other words, let the decision to leave at the end of period r be represented by:

$$I(C = r) \;=\; I(Y_r^* < Y_r) \prod_{j=1}^{r-1} I(Y_j^* \geq Y_j) \quad \text{for } r = 1, \ldots, R-1$$

whereas the decision to never leave be represented by $I(C = R) = 1 - \sum_{r=1}^{R-1} I(C = r)$.

The researcher observes C but not Y_r^*. This means that $Z_1 = (Y_{-1}, Y_0, Y_1, X_{-1}, X_0, X_1)'$,

[20] Estimation of regression coefficients in the case of attrition under some form of MAR in dynamic panel data models with fixed effects has been studied in, e.g., Abrevaya (2019).

$Z_2 = (Y_2, X_2)'$ and $Z_3 = (Y_3, X_3)'$ in our notation. So, the observables are $T_1 = Z_1$, $T_2 = (Z_1', Z_2')'$ and $T_3 = (Z_1', Z_2', Z_3')'$ for those with $C = 1$, $C = 2$ and $C = 3$ respectively.

Our distributional assumptions on the data generating process (DGP) are as follows. e_t and v_t are i.i.d. $N(0,1)$ for all t. $u_r := Y_r^* - Y_r$ is i.i.d. $N(0,7)$ for all r. We stipulate a rather large variance for u_r to abstract away from limited overlap. MAR in (1) is imposed by maintaining that $e_t, v_t, u_r, X_0, Y_{-1}, Y_0$ are mutually independent for all t, r. This results in roughly 62% of the individuals with $C = 1$, 23% with $C = 2$, and 15% with $C = 3$.

There are six different targets $[a,b] = [1,3], [1,1], [2,2], [3,3], [1,2]$ and $[2,3]$ that our theoretical results can accommodate for, and we have simulation results for all of them. For brevity, however, we will focus here on $[a,b] = [1,3], [1,1]$ and $[2,2,]$. ($[3,3]$ is the complete case and is trivial whereas the results for $[1,2]$ and $[2,3]$ are similar to those reported here.)

To define $\beta^0_{[a,b]}$, we take the moment function in (2) as $m(Z;\beta) = Y_3 - \beta$ and consider the three targets $[a,b] = [1,3], [1,1]$ and $[2,2]$ giving three parameters of interest. These target parameters are purposely defined similarly to the estimands in our empirical illustration.

We compute the "true value" of these target parameters numerically by generating data from the above DGP with sample size 10 million, estimating the mean of Y_3 for each sub-population, and then averaging each mean over 10,000 Monte Carlo trials. Consequently, the three target "true values" are: $\beta^0_{[1,3]} = 9.6162$, $\beta^0_{[1,1]} = 10.5232$ and $\beta^0_{[2,2]} = 8.9914$. As evident from Table 5, the error in this approximation is of a rather small order to seriously affect our subsequent analysis that is based on far smaller (than 10 million) sample size.

Target	Descriptive Statistics					
$[a,b]$ for β	Mean	Std	Median	IQR	Min	Max
$[1,3]$	9.6162	0.0022	9.6162	0.0029	9.6086	9.6249
$[1,1]$	10.5232	0.0027	10.5232	0.0037	10.5111	10.5329
$[2,2]$	8.9914	0.0044	8.9914	0.0060	8.9745	9.0084
$[3,3]$	6.8724	0.0050	6.8724	0.0067	6.8516	6.8924

Table 5: $\beta^0_{[a,b]}$ is approximated (column 2) for different target populations (column 1) based on averaging over 10,000 Monte Carlo trials the target-sample means obtained by using the same DGP and with sample size $n = 10$ million. Columns 3-7 list the standard deviation (Std), interquartile range (IQR), minimum (Min) and maximum (Max) of the estimator.

B.2 Simulation results

We compute our proposed estimator EFF following the description in Section 4. To estimate the nuisance parameters we use as working models the probit models for the conditional hazards and linear models for the conditional expectations. For each working model, we specify the index function as linear in the associated conditioning variables T_1, T_2, etc. and do not include interactions. The true conditional hazards $p^0(.)$'s but not the true conditional expectations $q^0(.)$'s are contained in their respective nuisance working models.

We report in Table 6 the simulation results based on these working models and 10,000 Monte Carlo trials, and for sample size n ranging from quite small to large.[21] We report: (i) Bias, the empirical mean bias; (ii) MC Std, the Monte Carlo standard deviation; (iii) AS Std, the average of the estimated standard error based on the asymptotic variance formula; and (iv) Size, the empirical size of the asymptotic 5% two-sided t test of $H_0 : \beta_{[a,b]} = \beta^0_{[a,b]}$.

Our proposed EFF performs very well in all these aspects (and others) and for all the target $\beta^0_{[a,b]}$ (including those unreported here) even when the sample size n is relatively small.

To put the performance of EFF in context, we also report the same properties of the IPW estimator from (14). IPW performs worse, often much worse, than EFF in every aspect.

First, consider empirical bias. The working parametric models contain the true conditional hazards, i.e., CH holds, and, therefore, IPW and EFF are both asymptotically unbiased. This shows for IPW in the simulation results if we focus on the relatively large samples. On the other hand, the empirical bias of EFF is quite small even in small samples.

Second, consider the variability of the IPW and EFF estimators. MC Std is of course infeasible in practice but is a better measure of the true variability. EFF seems to have much smaller MC Std than IPW. The same observation holds true for AS Std, which is the average of the estimated standard error, a feasible measure, from all the Monte Carlo trials.[22]

[21] $n = 200$ with $P(C = 3) \approx .15$ is small relative to the number of nuisance parameters; $n = 5000$ is not.

[22] We should however note that the observation that MC Std and AS Std are both smaller for EFF than IPW in our simulations is not theoretically promised. This is because: (i) although CH holds, the working models do not contain the true conditional expectations $q^0(.)$'s and hence EFF is not semiparametrically efficient, and (ii) we do not use the Cao et al. (2009)-modification of EFF that, in these cases of scalar

n	Target	Bias		MC Std		AS Std		Size	
	$[a,b]$	EFF	IPW	EFF	IPW	EFF	IPW	EFF	IPW
	[1,3]	-.043	-.229	.658	1.322	.587	1.006	8.6	15.4
200	[1,1]	-.053	-.330	.782	1.580	.728	1.212	7.1	16.9
	[2,2]	-.044	-.132	1.042	1.452	1.003	1.301	6.2	9.0
	[1,3]	-.030	-.147	.571	1.177	.523	.911	7.4	13.9
250	[1,1]	-.042	-.223	.680	1.429	.645	1.105	6.7	15.8
	[2,2]	-.021	-.072	.927	1.313	.897	1.169	6.2	7.8
	[1,3]	-.024	-.122	.520	1.044	.477	.822	7.8	12.6
300	[1,1]	-.032	-.189	.617	1.277	.586	1.004	6.4	14.4
	[2,2]	-.025	-.054	.845	1.172	.816	1.054	6.3	7.4
	[1,3]	-.021	-.102	.479	.944	.443	.764	7.2	11.7
350	[1,1]	-.033	-.160	.566	1.169	.544	.939	6.0	14.2
	[2,2]	-.007	-.040	.782	1.056	.757	.974	6.0	6.7
	[1,3]	-.014	-.079	.445	.882	.414	.714	6.8	10.8
400	[1,1]	-.019	-.125	.530	1.090	.508	.881	5.9	12.6
	[2,2]	-.012	-.033	.730	1.001	.709	.907	5.9	6.5
	[1,3]	-.020	-.062	.391	.782	.371	.643	6.8	10.1
500	[1,1]	-.024	-.093	.472	.988	.454	.800	5.9	11.6
	[2,2]	-.013	-.025	.643	.854	.633	.806	5.3	6.2
	[1,3]	-.004	-.033	.313	.615	.305	.530	5.4	8.6
750	[1,1]	-.008	-.055	.375	.785	.372	.667	4.9	9.9
	[2,2]	-.004	-.014	.522	.681	.518	.650	5.1	5.4
	[1,3]	-.002	-.005	.121	.222	.119	.213	5.5	6.3
5000	[1,1]	-.002	-.008	.145	.290	.145	.276	5.2	6.6
	[2,2]	-.004	-.005	.202	.248	.201	.245	5.1	5.1

Table 6: Results for EFF and IPW are reported based on 10,000 Monte Carlo trials and various sample sizes n. Bias stands for the empirical bias. MC Std and AS Std stands for the standard deviation based on Monte Carlo and the asymptotic variance formula respectively. Size stands for the empirical size of the asymptotic 5% two-sided t-test of $H_0 : \beta_{[a,b]} = \beta^0_{[a,b]}$.

We also note from Table 6 that the feasible measure AS Std resembles very well the infeasible but truer measure MC Std in the case of EFF. Interestingly, on the other hand, AS Std of IPW is much smaller than its MC Std. For practical purposes this means that the user's estimate of the standard error for IPW likely gives a misleadingly higher sense of precision especially in smaller samples. Theoretically, this indicates that the asymptotic approximation better resembles the small sample behavior of EFF than of IPW.

Finally, and extending the discussion of underestimated standard error and quality of asymptotic approximation, we consider Size. Size denotes the empirical size defined as the estimated probability of rejecting the truth by an asymptotic 5% two-sided t test for $H_0 : \beta_{[a,b]} = \beta^0_{[a,b]}$. We observe that Size is much closer to the nominal 5% level for EFF than it is for IPW. (IPW over-rejects the truth much more in small samples.[23]) This is doubly attractive for EFF in these simulations since, as anticipated from our observations on bias and variability, this shows that EFF's gain in precision over IPW comes with another advantage that EFF rejects the truth much less often than IPW, especially in small samples.

Now we move to the case where the nuisance parameters are nonparametrically estimated. The asymptotic variance of IPW estimators should decrease in such cases and, under suitable assumptions, can even reach the efficiency bound; see our Proposition 6(ii) in Section 3.3. Also see, e.g., Hirano et al. (2003), Wooldridge (2007), Chen et al. (2008), Graham (2011), Ackerberg et al. (2014), etc. in similar contexts and Newey (1994), Ackerberg et al. (2012), etc. more generally. We will pursue here this line of argument by obtaining the AS Std of the following three variants of the IPW estimator by enriching the original working model:

- IPW2: based on a working model that augments the original working model (for IPW in Table 6) with the squared terms but no interactions;

parameters of interest, would guarantee that the asymptotic variance of EFF is not larger than that of IPW if CH holds. Nevertheless, it is certainly a welcome observation that EFF delivers estimates that are much more precise than the IPW estimates. We have also noticed this in our other works with more than one level of missingness ($R > 2$). This discussion will need to be modified if the working models "promise" increased flexibility with sample size n; see Ackerberg et al. (2012); and we will do that later with the help of Table 7.

[23] Given Hahn and Liao (2021)'s result of the conservativeness of bootstrap standard error, this observation seems to justify that the anecdotally-common empirical practice of using bootstrap standard errors for IPW.

- IPW2in: based on a working model that augments the original working model (for IPW in Table 6) with the squared terms and all the first order interactions;
- IPW23: based on a working model that augments the original working model (for IPW in Table 6) with the squared and cubic terms but no interactions.

When the progressively richer working models used by these estimators are viewed as a function of sample size n, one would hope that these estimators' asymptotic variances computed as before would eventually converge to the efficiency bound; see, e.g., Newey (1994) and Ackerberg et al. (2012). We report in Table 7 the AS Std and MC Std of IPW2, IPW2in and IPW23 along with IPW and EFF for progressively large sample size. To abstract from: (i) the increased bias (unreported) in smaller samples that is not our focus but nevertheless important and well-studied (see, e.g., Chernozhukov et al. (2022), Rothe and Firpo (2019)) and (ii) more generally from any number of smaller sample issues (see, e.g., Sur and Candes (2019)), we even consider the extremely large sample size of $100,000$.

We also computed another variant IPW23in that is based on a working model that augments the original working model (for IPW in Table 6) with the squared and cubic terms and all first and second order interactions. However, we do not discuss IPW23in except in footnote 25 and omit it from Table 7 because it performs terribly except that when $n = 100,000$, its MC Std is slightly smaller than that of IPW23 (but still bigger, sometimes much bigger, than EFF) that in that instance is the best among the rest of the IPW variants.

We wish to discuss now several observations from Table 7.

First, continuing on the discussion of Table 6, the difference between MC Std and AS Std for each estimator ultimately vanishes with very large sample size ($n = 10,000$ or more).

Second, both MC Std and AS Std of IPW2, IPW2in and IPW23 are smaller than that of IPW for sample size $n = 5000$ and more. This ranking of variability is reassuring since the working models used by IPW2, IPW2in and IPW23 nest the model used by IPW.

Third, although the working models used by both IPW2in and IP23 nest the model used by IPW2, the variability of the former two, as measured by both MC Std and AS Std, seems

n	Target	MC Std					AS Std				
	$[a,b]$	EFF	IPW	IPW2	IPW2in	IPW23	EFF	IPW	IPW2	IPW2in	IPW23
1000	[1,3]	.275	.535	.461	.523	.452	.265	.465	.434	.549	.462
	[1,1]	.329	.687	.579	.633	.549	.322	.589	.539	.667	.569
	[2,2]	.448	.588	.553	.642	.577	.449	.562	.566	.707	.612
5000	[1,3]	.121	.222	.173	.187	.174	.119	.213	.162	.177	.178
	[1,1]	.145	.290	.227	.238	.221	.145	.276	.208	.224	.225
	[2,2]	.202	.248	.220	.235	.225	.201	.245	.220	.238	.235
10000	[1,3]	.085	.157	.113	.116	.114	.084	.152	.110	.114	.118
	[1,1]	.101	.205	.148	.151	.146	.102	.198	.142	.145	.149
	[2,2]	.144	.174	.150	.153	.154	.142	.173	.150	.157	.158
100000	[1,3]	.027	.049	.034	.033	.032	.027	.049	.034	.033	.032
	[1,1]	.033	.064	.044	.042	.040	.032	.064	.044	.042	.041
	[2,2]	.045	.055	.047	.047	.047	.045	.055	.046	.046	.046

Table 7: Standard deviations – MC Std and AS Std – based on Monte Carlo and the asymptotic variance formula respectively of EFF, IPW, IPW2, IPW2in and IPW23 are reported based on 10,000 Monte Carlo trials. The various versions of IPW differ in terms of the specification for the working model for the nuisance parameters – the conditional hazards and expectations, i.e., $p^0(.)$ and $q^0(.)$. In particular, EFF and IPW are those based on the original working model used in Table 6. IPW2 is based on a working model that augments the original working model with the squared terms but no interactions. IPW2in is based on a working model that augments the original working model with the squared terms and all the first order interactions. IPW23 is based on a working model that augments the original working model with the squared and cubic terms but no interactions.

to exceed that of IPW2 even for sample size as large as $n = 10,000$.

The above observations suggest that even in a simple framework such as ours, the sample size of $n = 10,000$ may not be large enough for the intuitions of the large sample theory of IPW to hold convincingly. Other basis functions could lead to a more encouraging picture. Nevertheless, our discussion based on the power series basis is practically relevant since power series resembles the common parametric specification of main variables and interactions used in empirical work and, therefore, it renders the transition from parametric to nonparametric specifications (by adding higher order terms) seamless and empirically palatable.

Fourth, the working models used by IPW2in and IPW23 do not nest each other and hence the ranking of the variability of IPW2in and IPW23 is theoretically unclear. The simulation results lead us to prefer IPW23. For this reason we use this working model in our empirical application (see footnote 14) where the sample size and the dimension of the covariates are comparable to those in the setup here. Some sort of formal regularization or variable selection could be useful, but that is beyond the scope of our current paper.

Finally, we observe from Table 7 that the variabilities, as measured by MC Std and AS Std, of IPW2, IPW2in and IPW23 are still worse, and sometimes much worse, than that of EFF even though EFF is based only on the original working model (as in Table 6).

Let us elaborate on this last observation because this also brings us back to one of our motivations behind extending the MAR analysis to sub-populations with multi-level missingness. To abstract away from any small sample issues that could have worked unfavorably for IPW2, IPW2in and IPW23 because of the large number of nuisance parameters involved in them, let us focus on an extreme case of very large sample size $n = 100,000$.[24]

Now the variabilities of IPW2, IPW2in and IPW23 come quite close to that of EFF for the target $\beta^0_{[2,2]}$. This is a case of only one level of missingness because $R = 3$ while $a = b = 2$; see footnote 9. One level of missingness is what has been considered in the cited references that showed nice properties of IPW based on nonparametric estimation of the conditional

[24]IPW2, IPW2in and IPW23 involve 19, 28 and 55 parameters respectively in their working models for $P(C = 2|C \geq 2, T_2)$ to be estimated based on approximately 38,000 observations ($C \geq 2$) when $n = 100,000$.

hazard (propensity score). Therefore, this closeness of variability and the realization of the promised benefit of nonparametrics is not surprising for the target $\beta^0_{[2,2]}$ when $R = 3$.

However, the variability of IPW2, IPW2in and IPW23 are still substantially larger than that of EFF for the target $\beta^0_{[1,1]}$ that is a case of two levels of missingness since $R = 3$ while $a = b = 1$. We observe the same for $\beta^0_{[1,3]}$ since $\beta^0_{[1,3]}$ is a weighted average involving $\beta^0_{[1,1]}$.[25]

We conclude by restating the three take away points. First, the promises of nonparametrics may not always hold even in very large samples. Second, it is indeed remarkable that the simple EFF estimator fared so well against the other estimators that were based on much richer working models. Third, apart from performing much better than IPW, EFF also performs well in all aspects in absolute terms even in samples of relatively small size.

C Bibliography

Abrevaya, J. (2019). Missing dependent variables in fixed-effects models. *Journal of Econometrics*, 211:151–165.

Ackerberg, D., Chen, X., and Hahn, J. (2012). A Practical Asymptotic Variance Estimator For Two-Step Semiparametric Estimators. *The Review of Economics and Statistics*, 94:481–498.

Ackerberg, D., Chen, X., Hahn, J., and Liao, Z. (2014). Asymptotic Efficiency of Semiparametric Two-step GMM. *Review of Economic Studies*, 81: 919–943.

Cao, W., Tsiatis, A., and Davidian, M. (2009). Improving Efficiency and Robustness of the Doubly Robust Estimator for a Population Mean with Incomplete Data. *Biometrika*, 96:723–734.

[25]As noted earlier when mentioning IPW23in, enriching the working model further uniformly is not useful here in reducing variability. One could selectively enrich the working model; e.g., enrich it much for $P(C = 1|C \geq 1, T_1)$ since T_1 is observed for all, but keep the model at the level of IPW23 for $P(C = 2|C \geq 2, T_2)$. However, T_1 should be nested in T_2 by the definition of monotonicity. Therefore, selective enrichments that lead to the working model for $P(C = 1|C \geq 1, T_1)$ not being nested in that for $P(C = 2|C \geq 2, T_2)$ are ultimately closer in spirit to imposing parametric restrictions on the MAR assumption. In that case, the target efficiency bounds need to be modified; see, e.g., Hahn (1998) and Chen et al. (2008).

Chaudhuri, S. (2020). On Efficiency Gains from Multiple Incomplete Subsamples. *Econometric Theory*, 36:488–525.

Chen, X., Hong, H., and Tarozzi, A. (2008). Semiparametric Efficiency in GMM Models with Auxiliary Data. *Annals of Statistics*, 36:808–843.

Chernozhukov, V., Escanciano, J.-C., Ichimura, H., Newey, W., and Robins, J. (2022). Locally Robust Ssemiparametric Estimation. *Econometrica*, 90: 1501–1535.

Graham, B. S. (2011). Efficiency Bounds for Missing Data Models with Semiparametric Restrictions. *Econometrica*, 79:437 – 452.

Hahn, J. (1998). On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects. *Econometrica*, 66:315–331.

Hahn, J. and Liao, Z. (2021). Bootstrap Standard Error Estimates and Inference. *Econometrica*, 89: 1963–1977.

Hirano, K., Imbens, G., and Ridder, G. (2003). Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Scores. *Econometrica*, 71:1161–1189.

Newey, W. (1994). The Asymptotic Variance of Semiparametric Estimators. *Econometrica*, 62:1349–1382.

Rothe, C. and Firpo, S. (2019). Properties of Doubly Robust Estimators when Nuisance Functions are Estimated Nonparametrically. *Econometric Theory*, 35: 1048–1087.

Sur, P. and Candes, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. *The Proceedings of the National Academy of Sciences*, 116: 14516–14525.

Wooldridge, J. M. (2007). Inverse Probability Weighted Estimation for General Missing Data Problems. *Journal of Econometrics*, 141(2):1281–1301.