The International Journal of Biostatistics

Volume 6, *Issue* 2 2010 *Article* 9

CAUSAL INFERENCE

Dynamic Regime Marginal Structural Mean Models for Estimation of Optimal Dynamic Treatment Regimes, Part II: Proofs of Results

Liliana Orellana* Andrea Rotnitzky† James M. Robins‡

*Instituto de Cálculo, Universidad de Buenos Aires, orellana@dm.uba.ar

†Universidad Torcuato Di Tella and Harvard School of Public Health, arotnitzky@utdt.edu

‡Harvard School of Public Health, robins@hsph.harvard.edu

Dynamic Regime Marginal Structural Mean Models for Estimation of Optimal Dynamic Treatment Regimes, Part II: Proofs of Results*

Liliana Orellana, Andrea Rotnitzky, and James M. Robins

Abstract

In this companion article to "Dynamic Regime Marginal Structural Mean Models for Estimation of Optimal Dynamic Treatment Regimes, Part I: Main Content" [Orellana, Rotnitzky and Robins (2010), IJB, Vol. 6, Iss. 2, Art. 7] we present (i) proofs of the claims in that paper, (ii) a proposal for the computation of a confidence set for the optimal index when this lies in a finite set, and (iii) an example to aid the interpretation of the positivity assumption.

KEYWORDS: dynamic treatment regime, double-robust, inverse probability weighted, marginal structural model, optimal treatment regime, causality

*This work was supported by NIH grant R01 GM48704.

1 Introduction

In this companion article to "Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes. Part I: Main Content" (Orellana, Rotnitzky and Robins, 2010) we present (i) proofs of the claims in that paper, (ii) a proposal for the computation of a confidence set for the optimal index when this lies in a finite set, and (iii) an example to aid the interpretation of the positivity assumption.

The notation, definitions and acronyms are the same as in the companion paper. Througout, we refer to the companion article as ORR-I.

2 Proof of Claims in ORR-I

2.1 Proof of Lemma 1

First note that the consistency assumption C implies that the event

$$\overline{O}_k = \overline{o}_k, \overline{A}_{k-1} = \overline{g}_{k-1}\left(\overline{o}_{k-1}\right)$$

is the same as the event

$$\overline{O}_k^g = \overline{o}_k, \overline{A}_{k-1} = \overline{g}_{k-1}\left(\overline{o}_{k-1}\right).$$

So, with the definitions

$$\underline{V}_{k,k+l} \equiv \left(V_{k+1}, ..., V_{k+l}\right), l > 0 \text{ and } \underline{V}_{k,k} \equiv \text{nil}$$

we obtain

$$\begin{aligned} & E\left[I_B\left(O, A\right) \underline{\omega}_{k-1,K}\left(\overline{O}_K, \overline{A}_K\right) | \overline{O}_k, \overline{A}_{k-1} = \overline{g}_{k-1}\left(\overline{O}_{k-1}\right)\right] \\ = & E\left[I_B\left(\left(\overline{O}_k^g, \underline{O}_{k,K+1}\right), \left(\overline{g}_{k-1}\left(\overline{O}_{k-1}^g\right), \underline{A}_{k-1,K}\right)\right)\right. \\ & \left.\underline{\omega}_{k-1,K}\left(\left(\overline{O}_k^g, \underline{O}_{k,K}\right), \left(\overline{g}_{k-1}\left(\overline{O}_{k-1}^g\right), \underline{A}_{k-1,K}\right)\right) | \overline{O}_k^g, \overline{A}_{k-1} = \overline{g}_{k-1}\left(\overline{O}_{k-1}^g\right)\right] \\ & \text{w.p.1.} \end{aligned}$$

Next, note that the fact that $\underline{\omega}_{k-1,K}\left(\left(\overline{O}_k^g, \underline{O}_{k,K}\right), \left(\overline{g}_{k-1}\left(\overline{O}_{k-1}^g\right), \underline{A}_{k-1,K}\right)\right) = 0$ unless $A_k = g_k\left(\overline{O}_k^g\right), A_{k+1} = g_{k+1}\left(\overline{O}_k^g, O_{k+1}\right), ..., A_K = g_{K+1}\left(\overline{O}_k^g, \underline{O}_{k,K}\right)$ implies that

$$\begin{aligned} & I_B\left(\left(\overline{O}_k^g, \underline{O}_{k,K+1}\right), \left(\overline{g}_{k-1}\left(\overline{O}_{k-1}^g\right), \underline{A}_{k-1,K}\right)\right) \times \\ & \underline{\omega}_{k-1,K}\left(\left(\overline{O}_k^g, \underline{O}_{k,K}\right), \left(\overline{g}_{k-1}\left(\overline{O}_{k-1}^g\right), \underline{A}_{k-1,K}\right)\right) \\ & = I_B\left(\overline{O}_{K+1}^g, \overline{g}_K\left(\overline{O}_K^g\right)\right) \underline{\omega}_{k-1,K}\left(\overline{O}_K^g, \left(\overline{g}_{k-1}\left(\overline{O}_{k-1}^g\right), \underline{A}_{k-1,K}\right)\right). \end{aligned}$$

Then, it follows from the second to last displayed equality that

$$\begin{aligned}
& E\left[I_B(O,A)\,\underline{\omega}_{k-1,K}\left(\overline{O}_K,\overline{A}_K\right)|\overline{O}_k,\overline{A}_{k-1}=\overline{g}_{k-1}\left(\overline{O}_{k-1}\right)\right] \\
= & E\left[I_B\left(\overline{O}^g_{K+1},\overline{g}_K\left(\overline{O}^g_K\right)\right)\underline{\omega}_{k-1,K}\left(\overline{O}^g_K,\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K}\right)\right)| \right.\\
& \left.\overline{O}^g_k,\overline{A}_{k-1}=\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right)\right] \\
= & E\left[E\left[\underline{\omega}_{k-1,K}\left(\overline{O}^g_K,\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K}\right)\right)|\overline{O}^g_{K+1},\overline{A}_{k-1}=\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right)\right]\times\right.\\
& \left.I_B\left(\overline{O}^g_{K+1},\overline{g}_K\left(\overline{O}^g_K\right)\right)|\overline{O}^g_k,\overline{A}_{k-1}=\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right)\right].
\end{aligned}$$

So, part 1 of the Lemma is proved if we show that

$$E\left[\underline{\omega}_{k-1,K}\left(\overline{O}^g_K,\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K}\right)\right)|\overline{O}^g_{K+1},\overline{A}_{k-1}=\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right)\right]=1. \quad (1)$$

Define for any $k=0,...,K$,

$$\underline{\omega}_{k,k}\left(\overline{O}^g_k,\left(\overline{g}_k\left(\overline{O}^g_k\right),\underline{A}_{k,k}\right)\right)\equiv 1.$$

To prove equality (1) first note that,

$$\begin{aligned}
& E\left[\underline{\omega}_{k-1,K}\left(\overline{O}^g_K,\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K}\right)\right)|\right.\\
& \left.\overline{O}^g_{K+1},\overline{A}_{k-1}=\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-1}\right] \\
= & \underline{\omega}_{k-1,K-1}\left(\overline{O}^g_{K-1},\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-1}\right)\right)\times\\
& E\left[\underline{\omega}_{K-1,K}\left(\overline{O}^g_K,\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K}\right)\right)\right|\\
& \left.\overline{O}^g_{K+1},\overline{A}_{k-1}=\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-1}\right] \\
= & \underline{\omega}_{k-1,K-1}\left(\overline{O}^g_{K-1},\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-1}\right)\right)\times\\
& E\left[\underline{\omega}_{K-1,K}\left(\overline{O}^g_K,\overline{A}_K\right)\right|\overline{O}^g_{K+1},\overline{A}_{K-1}=\overline{g}_{K-1}\left(\overline{O}^g_{K-1}\right)\right] \\
= & \underline{\omega}_{k-1,K-1}\left(\overline{O}^g_{K-1},\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-1}\right)\right)\times\\
& E\left[\left.\frac{I_{\left\{g_K\left(\overline{O}^g_K\right)\right\}}\left(A_K\right)}{\lambda_K\left(g_K\left(\overline{O}^g_K\right)|\overline{O}^g_K,\overline{g}_{K-1}\left(\overline{O}^g_{K-1}\right)\right)}\right|\overline{O}^g_{K+1},\overline{A}_{K-1}=\overline{g}_{K-1}\left(\overline{O}^g_{K-1}\right)\right] \\
= & \underline{\omega}_{k-1,K-1}\left(\overline{O}^g_{K-1},\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-1}\right)\right)\times\\
& \frac{E\left[I_{\left\{g_K\left(\overline{O}^g_K\right)\right\}}\left(A_K\right)\right|\overline{O}^g_{K+1},\overline{A}_{K-1}=\overline{g}_{K-1}\left(\overline{O}^g_{K-1}\right)\right]}{\lambda_K\left(g_K\left(\overline{O}^g_K\right)|\overline{O}^g_K,\overline{g}_{K-1}\left(\overline{O}^g_{K-1}\right)\right)} \\
= & \underline{\omega}_{k-1,K-1}\left(\overline{O}^g_{K-1},\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-1}\right)\right)\times\\
& \frac{\mathbb{P}\left[A_K=g_K\left(\overline{O}^g_K\right)\right|\overline{O}^g_{K+1},\overline{A}_{K-1}=\overline{g}_{K-1}\left(\overline{O}^g_{K-1}\right)\right]}{\lambda_K\left(g_K\left(\overline{O}^g_K\right)|\overline{O}^g_K,\overline{g}_{K-1}\left(\overline{O}^g_{K-1}\right)\right)}
\end{aligned}$$

$$
\begin{aligned}
&= \underline{\omega}_{k-1,K-1}\left(\overline{O}^g_{K-1},\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-1}\right)\right)\times \\
&\quad \frac{\mathbb{P}\left[A_K = g_K\left(\overline{O}^g_K\right)\middle|\,\overline{O}^g_K,\overline{A}_{K-1}=\overline{g}_{K-1}\left(\overline{O}^g_{K-1}\right)\right]}{\lambda_K\left(g_K\left(\overline{O}^g_K\right)|\overline{O}^g_K,\overline{g}_{K-1}\left(\overline{O}^g_{K-1}\right)\right)} \\
&= \underline{\omega}_{k-1,K-1}\left(\overline{O}^g_{K-1},\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-1}\right)\right)
\end{aligned}
$$

where the second to last equality follows because given $\overline{O}^g_K$ and $\overline{A}_{K-1} = \overline{g}_{K-1}\left(\overline{O}^g_{K-1}\right)$, O^g_{K+1} is a fixed, i.e. non-random function of $\mathcal{O}$ and consequently by the sequential randomization assumption, O^g_{K+1} is conditionally independent of A_K given $\overline{O}^g_K$ and $\overline{A}_{K-1} = \overline{g}_{K-1}\left(\overline{O}^g_{K-1}\right)$. The last equality follows by the definition of $\lambda_K(\cdot|\cdot,\cdot)$.

We thus arrive at

$$
\begin{aligned}
& E\left[\underline{\omega}_{k-1,K}\left(\overline{O}^g_K,\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K}\right)\right)\middle|\,\overline{O}^g_{K+1},\overline{A}_{k-1}=\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right)\right] \\
=\; & E\left\{E\left[\underline{\omega}_{k-1,K}\left(\overline{O}^g_K,\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K}\right)\right)\middle|\right.\right. \\
& \left.\left.\overline{O}^g_{K+1},\overline{A}_{k-1}=\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-1}\right]\middle|\,\overline{O}^g_{K+1},\overline{A}_{k-1}=\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right)\right\} \\
=\; & E\left\{\underline{\omega}_{k-1,K-1}\left(\overline{O}^g_{K-1},\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-1}\right)\right)\middle|\right. \\
& \left.\overline{O}^g_{K+1},\overline{A}_{k-1}=\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right)\right\}
\end{aligned}
$$

This proves the result for the case $k = K$. If $k < K-1$, we analyze the conditional expectation of the last equality in a similar fashion. Specifically, following the same steps as in the long sequence of equalities in the second to last display we arrive at

$$
\begin{aligned}
& E\left\{\underline{\omega}_{k-1,K-1}\left(\overline{O}^g_{K-1},\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-1}\right)\right)\middle|\right. \\
& \left.\overline{O}^g_{K+1},\overline{A}_{k-1}=\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-2}\right\} \\
=\; & \underline{\omega}_{k-1,K-2}\left(\overline{O}^g_{K-2},\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-2}\right)\right)\times \\
& \times\frac{\mathbb{P}\left[A_{K-1}=g_{K-1}\left(\overline{O}^g_{K-1}\right)\middle|\,\overline{O}^g_{K+1},\overline{A}_{K-2}=\overline{g}_{K-2}\left(\overline{O}^g_{K-2}\right)\right]}{\lambda_K\left(g_{K-1}\left(\overline{O}^g_{K-1}\right)|\overline{O}^g_{K-1},\overline{g}_{K-2}\left(\overline{O}^g_{K-2}\right)\right)} \\
=\; & \underline{\omega}_{k-1,K-2}\left(\overline{O}^g_{K-2},\left(\overline{g}_{k-1}\left(\overline{O}^g_{k-1}\right),\underline{A}_{k-1,K-2}\right)\right)
\end{aligned}
$$

the last equality follows once again from the sequential randomization assumption. This is so because given $\overline{O}^g_{K-1}$ and $\overline{A}_{K-2} = \overline{g}_{K-2}\left(\overline{O}^g_{K-2}\right)$, $\overline{O}^g_K$ and $\overline{O}^g_{K+1}$ are fixed, i.e. deterministic, functions of $\mathcal{O}$ and the SR assumption ensures then that $\overline{O}^g_K$ and $\overline{O}^g_{K+1}$ are conditionally independent of A_{K-1} given $\overline{O}^g_{K-1}$ and $\overline{A}_{K-2} = \overline{g}_{K-2}\left(\overline{O}^g_{K-2}\right)$.

Equality (1) is thus shown by continuing in this fashion recursively for $K-2, K-3, ..., K-l$ until l such that $K-l = k-1$.

To show Part 2 of the Lemma, note that specializing part 1 to the case $k = 0$, we obtain

$$E\left[I_B\left(O^g, A^g\right)|O_0\right] = E\left[I_B\left(O, A\right)\omega_K\left(\overline{O}_K, \overline{A}_K\right)|O_0\right].$$

Thus, taking expectations on both sides of the equality in the last display we obtain

$$E\left[I_B\left(O^g, A^g\right)\right] = E\left[I_B\left(O, A\right)\omega_K\left(\overline{O}_K, \overline{A}_K\right)\right].$$

This shows part 2 because B is an arbitrary Borel set.

2.2 Proof of the Assertions in Section 3.2, ORR-I

2.2.1 Proof of Item (a)

Lemma 1, part 2 implies that the densities p_g^{marg} factors as

$$p_g^{\text{marg}}\left(o, a\right) = \prod_{j=0}^{K} I_{\{g_j(\overline{o}_j)\}}\left(a_j\right) \prod_{j=1}^{K+1} p^{\text{marg}}\left(o_j|\overline{o}_{j-1}, \overline{a}_{j-1}\right) p^{\text{marg}}\left(o_0\right).$$

In particular, the event $\left\{\overline{A}_{k-1}^g = \overline{g}_{k-1}\left(\overline{O}_{k-1}^g\right)\right\}$ has probability 1. Consequently,

$$\begin{aligned} p_g^{\text{marg}}\left(o, a|\overline{o}_k\right) &= \prod_{j=0}^{K} I_{\{g_j(\overline{o}_j)\}}\left(a_j\right) p_g^{\text{marg}}\left(o, a|\overline{o}_k, \overline{a}_k = \overline{g}_k\left(\overline{o}_k\right)\right) \\ &= \prod_{j=0}^{K} I_{\{g_j(\overline{o}_j)\}}\left(a_j\right) \prod_{j=k+1}^{K+1} p^{\text{marg}}\left(o_j|\overline{o}_{j-1}, \overline{a}_{j-1} = \overline{g}_{j-1}\left(\overline{o}_{j-1}\right)\right). \end{aligned}$$

Therefore,

$$\begin{aligned} & E\left\{u\left(O^g, A^g\right)|\overline{O}_k^g = \overline{o}_k\right\} = \qquad (2) \\ &= \sum_{\substack{a_l \in \mathcal{A}_l \\ l=0,\ldots,K}} \int u\left(o, a\right) \prod_{j=0}^{K} I_{\{g_j(\overline{o}_j)\}}\left(a_j\right) \prod_{j=k+1}^{K+1} dP^{\text{marg}}_{O_j|\overline{O}_{j-1}, \overline{A}_{j-1}}\left(o_j|\overline{o}_{j-1}, \overline{g}_{j-1}\left(\overline{o}_{j-1}\right)\right) \\ &= \int u\left(o, a\right)\left[\sum_{\substack{a_l \in \mathcal{A}_l \\ l=0,\ldots,K}} \prod_{j=0}^{K} I_{\{g_j(\overline{o}_j)\}}\left(a_j\right)\right] \prod_{j=k+1}^{K+1} dP^{\text{marg}}_{O_j|\overline{O}_{j-1}, \overline{A}_{j-1}}\left(o_j|\overline{o}_{j-1}, \overline{g}_{j-1}\left(\overline{o}_{j-1}\right)\right) \\ &= \int u\left(o, a\right) \prod_{j=k+1}^{K+1} dP^{\text{marg}}_{O_j|\overline{O}_{j-1}, \overline{A}_{j-1}}\left(o_j|\overline{o}_{j-1}, \overline{g}_{j-1}\left(\overline{o}_{j-1}\right)\right) \end{aligned}$$

DOI: 10.2202/1557-4679.1242

$$
\begin{aligned}
&= \int \prod_{j=k+1}^{K} dP^{\text{marg}}_{O_j|\overline{O}_{j-1},\overline{A}_{j-1}} \left(o_j|\overline{o}_{j-1}, \overline{g}_{j-1}\left(\overline{o}_{j-1}\right)\right) \times \\
&\quad \left[\int u\left(o,a\right) dP^{\text{marg}}_{O_{K+1}|\overline{O}_K,\overline{A}_K} \left(o_{K+1}|\overline{o}_K, \overline{a}_K = \overline{g}_K\left(\overline{o}_K\right)\right)\right] \\
&= \int \left[\prod_{j=k+1}^{K} dP^{\text{marg}}_{O_j|\overline{O}_{j-1},\overline{A}_{j-1}} \left(o_j|\overline{o}_{j-1}, \overline{g}_{j-1}\left(\overline{o}_{j-1}\right)\right)\right] \phi_{K+1}\left(\overline{o}_K\right) \\
&= \int \left[\prod_{j=k+1}^{K-1} dP^{\text{marg}}_{O_j|\overline{O}_{j-1},\overline{A}_{j-1}} \left(o_j|\overline{o}_{j-1}, \overline{g}_{j-1}\left(\overline{o}_{j-1}\right)\right)\right] \times \\
&\quad \left[\int \phi_{K+1}\left(\overline{o}_K\right) dP^{\text{marg}}_{O_K|\overline{O}_{K-1},\overline{A}_{K-1}} \left(o_K|\overline{o}_{K-1}, \overline{g}_{K-1}\left(\overline{o}_{K-1}\right)\right)\right] \\
&= \int \prod_{j=k+1}^{K-1} dP^{\text{marg}}_{O_j|\overline{O}_{j-1},\overline{A}_{j-1}} \left(o_j|\overline{o}_{j-1}, \overline{g}_{j-1}\left(\overline{o}_{j-1}\right)\right) \phi_{K+1}\left(\overline{o}_K\right) \\
&= \ldots = \phi_{k+1}\left(\overline{o}_k\right).
\end{aligned}
$$

2.2.2 Proof of Item (b)

Lemma 1, part 1 implies that

$$
\begin{aligned}
&E\left[u\left(O,A\right) \underline{\omega}_{k-1,K}\left(\overline{O}_K, \overline{A}_K\right) |\overline{O}_k, \overline{A}_{k-1} = \overline{g}_{k-1}\left(\overline{O}_{k-1}\right)\right] = \\
&= E\left[u\left(O^g, A^g\right) |\overline{O}_k, \overline{A}_{k-1} = \overline{g}_{k-1}\left(\overline{O}_{k-1}\right)\right].
\end{aligned}
$$

The left hand side of this equality is equal to

$$
\sum_{\substack{a_k \in \mathcal{A}_k \\ k=k,\ldots,K}} \int u\left(o,a\right) \underline{\omega}_{k-1,K}\left(\overline{o}_K, \overline{a}_K\right) \prod_{j=k}^{K} \lambda_j\left(a_j|\overline{o}_j, \overline{a}_{j-1}\right) \prod_{j=k+1}^{K+1} dP^{\text{marg}}\left(o_j|\overline{o}_{j-1}, \overline{a}_{j-1}\right)
$$

and this coincides with the right hand side of (2) which, as we have just argued, is equal to $\phi_{k+1}\left(\overline{o}_k\right)$.

2.3 Proof of Lemma 2 in ORR-I

Let X be the identity random element on $(\mathcal{X}, \mathcal{A})$ and let $E_{P^{\text{marg}} \times P_X}(\cdot)$ stand for the expectation operation computed under the product law $P^{\text{marg}} \times P_X$ for the random vector (O, A, X). Then the restriction stated in 2) is equivalent to

$$
E_{P^{\text{marg}} \times P_X}\left[b\left(X,Z\right) \omega_K\left(\overline{O}_K, \overline{A}_K\right) \left\{u\left(O,A\right) - h_{\text{par}}\left(X,Z;\beta^*\right)\right\}\right] = 0 \text{ for all } b \qquad (3)
$$

and the restriction stated in 3) is equivalent to

$$E_{P^{\text{marg}}\times P_X}\left[\{b(X,Z)-E_{P\times P_X}[b(X,Z)|Z]\}\times \omega_K\left(\overline{O}_K,\overline{A}_K\right)\{u(O,A)-h_{\text{sem}}(X,Z;\beta^*)\}\right]=0 \text{ for all } b. \tag{4}$$

To show 2) let $d(O,A,X)\equiv\omega_K\left(\overline{O}_K,\overline{A}_K\right)\{u(O,A)-h_{\text{par}}(X,Z;\beta^*)\}$.

(ORR-I, (14)) $\Rightarrow$ (3).

$$\begin{aligned} E_{P^{\text{marg}}\times P_X}[b(X,Z)d(O,A,X)] &= E_{P^{\text{marg}}\times P_X}[b(X,Z)E_{P\times P_X}[d(O,A,X)|X,Z]] \\ &= 0 \end{aligned}$$

where the last equality follows because $E_{P^{\text{marg}}\times P_X}[d(O,A,X)|X=x,Z]=E_{P^{\text{marg}}}[d(O,A,x)|Z]$ by independence of (O,A) with X under the law $P^{\text{marg}}\times P_X$ and, by assumption, $E_{P^{\text{marg}}}[d(O,A,x)|Z]=0$ μ-a.e.(x) and hence $E_{P^{\text{marg}}}[d(O,A,x)|Z]$ because P_X and μ are mutually absolute continuous.

(3) $\Rightarrow$ (ORR-I, (14)). Define $b^*(X,Z)=E_{P^{\text{marg}}\times P_X}[d(O,A,X)|X,Z]$. Then,

$$0=E_{P^{\text{marg}}\times P_X}[b^*(X,Z)d(O,A,X)]=E_{P^{\text{marg}}\times P_X}\left[E_{P\times P_X}[d(O,A,X)|X,Z]^2\right]$$

consequently, $E_{P^{\text{marg}}\times P_X}[d(O,A,X)|X,Z]=0$ with $P^{\text{marg}}\times P_X$ prob. 1 which is equivalent to (ORR-I, (14)) because P_X is mutually absolutely continuous with μ.

To show 3) redefine $d(O,A,X)$ as $\omega_K\left(\overline{O_K},\overline{A_K}\right)\{u(O,A)-h_{\text{sem}}(X,Z;\beta^*)\}$.

(ORR-I, (15)) $\Rightarrow$ (4)

$$\begin{aligned} &E_{P^{\text{marg}}\times P_X}[\{b(X,Z)-E_{P^{\text{marg}}\times P_X}[b(X,Z)|Z]\}d(O,A,X)] \\ &=E_{P^{\text{marg}}\times P_X}[\{b(X,Z)-E_{P\times P_X}[b(X,Z)|Z]\}E_{P^{\text{marg}}\times P_X}\{d(O,A,X)|X,Z\}] \\ &=E_{P^{\text{marg}}\times P_X}[\{b(X,Z)-E_{P^{\text{marg}}\times P_X}[b(X,Z)|Z]\}q(Z)]=0 \end{aligned}$$

where the third equality follows because $E_{P^{\text{marg}}\times P_X}\{d(O,A,X)|X=x,Z\}=E_{P^{\text{marg}}}\{d(O,A,x)|Z\}$ and $E_{P^{\text{marg}}}\{d(O,A,x)|Z\}=q(Z)$ μ-a.e.(x) and hence P_X-a.e.(x) by absolute continuity.

(4) $\Rightarrow$ (ORR-I, (15)). Define $b^*(X,Z)=E_{P\times P_X}[d(O,A,X)|X,Z]$. Then,

$$\begin{aligned} 0&=E_{P^{\text{marg}}\times P_X}[\{b^*(X,Z)-E_{P^{\text{marg}}\times P_X}[b^*(X,Z)|Z]\}d(O,A,X)] \\ &=E_{P^{\text{marg}}\times P_X}[\{b^*(X,Z)-E_{P^{\text{marg}}\times P_X}[b^*(X,Z)|Z]\}b^*(X,Z)] \\ &=E_{P^{\text{marg}}\times P_X}\left[\{b^*(X,Z)-E_{P^{\text{marg}}\times P_X}[b^*(X,Z)|Z]\}^2\right]. \end{aligned}$$

Consequently, $b^*(X,Z)=E_{P^{\text{marg}}\times P_X}[b^*(X,Z)|Z]\equiv q(Z)\ P_X-a.e.(X)$ and hence $\mu_X-a.e.(X)$ by absolute continuity. The result follows because $b^*(x,Z)=E_{P^{\text{marg}}\times P_X}[d(O,A,X)|X=x,Z]=E_{P^{\text{marg}}}[d(O,A,X)|Z]$.

DOI: 10.2202/1557-4679.1242

2.4 Derivation of Some Formulas in Section 5.3, ORR-I

2.4.1 Derivation of Formula (26) in ORR-I

Any element

$$\sum_{k=0}^{K} \left\{ d_k \left(\overline{O}_k, \overline{A}_k \right) - E \left[d_k \left(\overline{O}_k, \overline{A}_k \right) | \overline{O}_k, \overline{A}_{k-1} \right] \right\}$$

of the set Λ is the sum of $K+1$ uncorrelated terms because for any l, j such that $0 \leq l < l + j \leq K + 1$,

$$\begin{aligned}
&E \left[\left\{ d_{l+j} \left(\overline{O}_{l+j}, \overline{A}_{l+j} \right) - E \left[d_{l+j} \left(\overline{O}_{l+j}, \overline{A}_{l+j} \right) | \overline{O}_{l+j}, \overline{A}_{l+j-1} \right] \right\} \times \right. \\
&\qquad \left. \left\{ d_l \left(\overline{O}_l, \overline{A}_l \right) - E \left[d_l \left(\overline{O}_l, \overline{A}_l \right) | \overline{O}_l, \overline{A}_{l-1} \right] \right\} \right] \\
&= E \left[E \left[\left\{ d_{l+j} \left(\overline{O_{l+j}}, \overline{A_{l+j}} \right) - E \left[d_{l+j} \left(\overline{O_{l+j}}, \overline{A_{l+j}} \right) | \overline{O_{l+j}}, \overline{A_{l+j-1}} \right] \right\} \middle| \overline{O_{l+j}}, \overline{A_{l+j-1}} \right] \times \right. \\
&\qquad \left. \left\{ d_l \left(\overline{O}_l, \overline{A}_l \right) - E \left[d_l \left(\overline{O}_l, \overline{A}_l \right) | \overline{O}_l, \overline{A}_{l-1} \right] \right\} \right] \\
&= E \left[0 \times \left\{ d_l \left(\overline{O}_l, \overline{A}_l \right) - E \left[d_l \left(\overline{O}_l, \overline{A}_l \right) | \overline{O}_l, \overline{A}_{l-1} \right] \right\} \right] = 0.
\end{aligned}$$

Thus, Λ is equal to $\Lambda_0 \oplus \Lambda_1 \oplus \cdots \oplus \Lambda_K$ where

$$\Lambda_k \equiv \left\{ d_k \left(\overline{O}_k, \overline{A}_k \right) - E \left[d_k \left(\overline{O}_k, \overline{A}_k \right) | \overline{O}_k, \overline{A}_{k-1} \right] : d_k \text{ arbitrary scalar function} \right\}$$

and $\oplus$ stands for the direct sum operator. Then,

$$\Pi \left[Q | \Lambda \right] = \sum_{k=0}^{K} \Pi \left[Q | \Lambda_k \right]$$

and it can be easily checked that $\Pi \left[Q | \Lambda_k \right] = E \left(Q | \overline{O}_k, \overline{A}_k \right) - E \left[Q | \overline{O}_k, \overline{A}_{k-1} \right]$.

2.4.2 Derivation of Formula (27) in ORR-I

Applying formula (26, in ORR-I) we obtain

$$\Pi \left[S_{\cdot} \left(\beta, \gamma^*, b \right) | \Lambda \right] = \sum_{k=0}^{K} \left\{ E \left[S_{\cdot} \left(\beta, \gamma^*, b \right) | \overline{O}_k, \overline{A}_k \right] - E \left[S_{\cdot} \left(\beta, \gamma^*, b \right) | \overline{O}_k, \overline{A}_{k-1} \right] \right\}.$$

So, for $k = 0, ..., K$,

$$d^{b}_{\cdot,opt,k} \left(\overline{O}_k, \overline{A}_k \right) = E \left[S_{\cdot} \left(\beta, \gamma^*, b \right) | \overline{O}_k, \overline{A}_k \right].$$

But,

$$
\begin{aligned}
& E\left[S.\left(\beta,\gamma^{*},b\right)|\overline{O}_k,\overline{A}_k\right] \\
= & \int_{\mathcal{X}_{pos}} b.\left(x,z\right)E\left[\omega^x_K\left(\overline{O}_K,\overline{A}_K\right)\left\{u\left(O,A\right)-h.\left(x,Z;\beta\right)\right\}\Big|\,\overline{O}_k,\overline{A}_k\right]dP_X\left(x\right) \\
= & \int_{\mathcal{X}_{pos}} b.\left(x,z\right)\omega^x_k\left(\overline{O}_k,\overline{A}_k\right)\times \\
& \times E\left[\underline{\omega}^x_{k,K}\left(\overline{O}_K,\overline{A}_K\right)\left\{u\left(O,A\right)-h.\left(x,Z;\beta\right)\right\}\Big|\,\overline{O}_k,\overline{A}_k\right]dP_X\left(x\right) \\
= & \int_{\mathcal{X}_{pos}} b.\left(x,z\right)\omega^x_k\left(\overline{O}_k,\overline{A}_k\right)\times \\
& \times E\left[\underline{\omega}^x_{k,K}\left(\overline{O}_K,\overline{A}_K\right)\left\{u\left(O,A\right)-h.\left(x,Z;\beta\right)\right\}\Big|\,\overline{O}_k,\overline{A}_k=g_x\left(\overline{O}_k\right)\right]dP_X\left(x\right).
\end{aligned}
$$

So formula ((27), ORR-I) is proved if we show that

$$
\begin{gathered}
E\left[\underline{\omega}^x_{k,K}\left(\overline{O}_K,\overline{A}_K\right)\left\{u\left(O,A\right)-h.\left(x,Z;\beta\right)\right\}\Big|\,\overline{O}_k,\overline{A}_k=g_x\left(\overline{O}_k\right)\right]= \\
\left\{\phi^x_{k+1}\left(\overline{O}_k\right)-h.\left(x,Z;\beta\right)\right\}.
\end{gathered}
\tag{5}
$$

This follows immediately from the preceding proof of Result (b) of Section 3.2. Specifically, it was shown there that

$$
E\left[\underline{\omega}^x_{k,K}\left(\overline{O}_K,\overline{A}_K\right)\left\{u\left(O,A\right)-h.\left(x,Z;\beta\right)\right\}\Big|\,\overline{O}_{k+1},\overline{A}_k=g_x\left(\overline{O}_k\right)\right]=\phi_{k+2}\left(\overline{O}_{k+1}\right).
$$

Consequently, the left hand side of (5) is equal to

$$
\begin{aligned}
& E\left[E\left[\underline{\omega}^x_{k,K}\left(\overline{O}_K,\overline{A}_K\right)\left\{u\left(O,A\right)-h.\left(x,Z;\beta\right)\right\}\Big|\,\overline{O}_{k+1},\overline{A}_k=g_x\left(\overline{O}_k\right)\right]\,|\right. \\
& \left.\overline{O}_k,\overline{A}_k=g_x\left(\overline{O}_k\right)\right] \\
= & E\left[\phi^x_{k+2}\left(\overline{O}_{k+1}\right)|\overline{O}_k,\overline{A}_k=g_x\left(\overline{O}_k\right)\right] \\
& -h.\left(x,Z;\beta\right)E\left[\underline{\omega}^x_{k,K}\left(\overline{O}_K,\overline{A}_K\right)|\overline{O}_k,\overline{A}_k=g_x\left(\overline{O}_k\right)\right] \\
= & \phi^x_{k+1}\left(\overline{O}_k\right)-h.\left(x,Z;\beta\right)
\end{aligned}
$$

where the last equality follows by the definition of $\phi^x_{k+1}\left(\overline{O}_k\right)$ and the fact that $E\left[\underline{\omega}^x_{k,K}\left(\overline{O}_K,\overline{A}_K\right)|\overline{O}_k,\overline{A}_k=g_x\left(\overline{O}_k\right)\right]=1$ (as this is just the function $\phi^x_{k+1}\left(\overline{O}_k\right)$ resulting from applying the integration to the utility $u\left(O,A\right)=1$).

2.4.3 Derivation of Formula (31) in ORR-I

It suffices to show that $S_{aug}\left(\gamma,d^b_{\cdot,\beta,\gamma,\tau,opt}\right)=\sum_{k=0}^{K}\int_{\mathcal{X}_{pos}}b\left(x,Z\right)M_k\left(x;\beta,\gamma,\tau\right)dP_X\left(x\right)$ where

$$
M_k\left(x;\beta,\gamma,\tau\right)\equiv\left\{\omega^x_k\left(\gamma\right)-\omega^x_{k-1}\left(\gamma\right)\right\}\left\{\phi^x_{k+1}\left(\overline{O}_k;\tau\right)-h.\left(x,Z;\beta\right)\right\}.
$$

But by definition

$$S_{aug}\left(\gamma, d^{b}_{\cdot,\beta,\gamma,\tau,opt}\right) =$$
$$= \sum_{k=0}^{K} \left\{ d^{b}_{\cdot,\beta,\gamma,\tau,opt,k}\left(\overline{O}_k, \overline{A}_k\right) - E_\gamma\left[d^{b}_{\cdot,\beta,\gamma,\tau,opt,k}\left(\overline{O}_k, \overline{A}_k\right)|\overline{O}_k, \overline{A}_{k-1}\right]\right\}$$
$$= \sum_{k=0}^{K} \left\{ \int_{\mathcal{X}_{pos}} b(x,Z)\,\omega_k^x(\gamma)\left\{\phi^x_{k+1}\left(\overline{O}_k;\tau\right) - h_\cdot(x,Z;\beta)\right\} dP_X(x) - \right.$$
$$\left. - E\left[\int_{\mathcal{X}_{pos}} b(x,Z)\,\omega_k^x(\gamma)\left\{\phi^x_{k+1}\left(\overline{O}_k;\tau\right) - h_\cdot(x,Z;\beta)\right\} dP_X(x)\,|\overline{O}_k, \overline{A}_{k-1}\right]\right\}$$
$$= \sum_{k=0}^{K} \int_{\mathcal{X}_{pos}} b(x,Z)\left\{\omega_k^x(\gamma) - E_\gamma\left[\omega_k^x(\gamma)\,|\overline{O}_k, \overline{A}_{k-1}\right]\right\}\times$$
$$\left\{\phi^x_{k+1}\left(\overline{O}_k;\tau\right) - h_\cdot(x,Z;\beta)\right\} dP_X(x)$$
$$= \sum_{k=0}^{K} \int_{\mathcal{X}_{pos}} b(x,Z)\left\{\omega_k^x(\gamma) - \omega_{k-1}^x(\gamma)\right\}\left\{\phi^x_{k+1}\left(\overline{O}_k;\tau\right) - h_\cdot(x,Z;\beta)\right\} dP_X(x)$$

where the last equality follows because

$$E_\gamma\left[\omega_k^x(\gamma)\,|\overline{O}_k, \overline{A}_{k-1}\right] =$$
$$= \omega_{k-1}^x(\gamma)\, E_\gamma\left[\left.\frac{I_{\left\{g_{x,k}\left(\overline{O}_k\right)\right\}}(A_k)}{\lambda_k\left(A_k|\overline{O}_k, \overline{A}_{k-1}\right)}\right|\overline{O}_k, \overline{A}_{k-1}\right]$$
$$= \omega_{k-1}^x(\gamma)\,\frac{E_\gamma\left[I_{\left\{g_{x,k}\left(\overline{O}_k\right)\right\}}(A_k)\Big|\,\overline{O}_k, \overline{A}_{k-1}\right]}{\lambda_k\left(g_{x,k}\left(\overline{O}_k\right)|\overline{O}_k, \overline{A}_{k-1}\right)} = \omega_{k-1}^x(\gamma)\,.$$

2.5 Proof that $b_{\cdot,opt}$ is Optimal

Write for short, $\widehat{\beta}_\cdot(b) \equiv \widehat{\beta}_\cdot\left(b, \widehat{d}_{\cdot,opt}\right)$,

$Q_{par}(b) \equiv \int_{\mathcal{X}_{os}} b(x,Z)\, Q_{par}\left(x;\beta^\dagger,\gamma^\dagger,\tau^\dagger\right) dP_X(x)$ and

$Q_{sem}(b) \equiv \int_{\mathcal{X}_{pos}} \left\{b(x,Z) - \overline{b}(Z)\right\} Q_{sem}\left(x;\beta^\dagger,\gamma^\dagger,\tau^\dagger\right) dP_X(x)$.

We will show that $J_\cdot(b) = E\left\{Q_\cdot(b)\, Q_\cdot(b_{\cdot,opt})'\right\}$ for $\cdot = par$ and $\cdot = sem$. When either model (16, ORR-I) or (29, ORR-I) are correct, $\beta^* = \beta^\dagger$. Consequently, for $\cdot = par$ we have that $J_{par}(b)$ is equal to

$$
\begin{aligned}
&- E\left\{\int_{\mathcal{X}_{pos}} b\left(x, Z\right) \frac{\partial}{\partial \beta} h_{\text{par}}\left(x, Z; \beta\right)\Big|_{\beta^{\dagger}} dP_X\left(x\right)\right\} \\
&= E\left[\int_{\mathcal{X}_{pos}} b\left(x, Z\right) dP_X\left(x\right) \times \right. \\
&\left. \times \left\{\int_{\mathcal{X}_{pos}} b_{par,opt}\left(x, Z\right) E\left\{Q_{par}\left(x; \beta^{\dagger}, \gamma^{\dagger}, \tau^{\dagger}\right) Q_{par}\left(\widetilde{x}; \beta^{\dagger}, \gamma^{\dagger}, \tau^{\dagger}\right)' |Z\right\} dP_X\left(\widetilde{x}\right)\right\}\right] \\
&= E\left[\left\{\int_{\mathcal{X}_{pos}} b\left(x, Z\right) Q_{par}\left(x; \beta^{\dagger}, \gamma^{\dagger}, \tau^{\dagger}\right) dP_X\left(x\right)\right\} \times \right. \\
&\left. \times \left\{\int_{\mathcal{X}_{pos}} b_{par,opt}\left(x, Z\right) Q_{par}\left(\widetilde{x}; \beta^{\dagger}, \gamma^{\dagger}, \tau^{\dagger}\right)' dP_X\left(\widetilde{x}\right)\right\}\right] \\
&= E\left\{Q_{par}\left(b\right) Q_{par}\left(b_{par,opt}\right)'\right\}.
\end{aligned}
$$

For $\cdot = sem$ and with the definitions $\widetilde{b}\left(x, Z\right) \equiv b\left(x, Z\right) - \overline{b}\left(Z\right)$ and $\widetilde{Q}_{sem}\left(\widetilde{x}; \beta^{\dagger}, \gamma^{\dagger}, \tau^{\dagger}\right) \equiv Q_{sem}\left(\widetilde{x}; \beta^{\dagger}, \gamma^{\dagger}, \tau^{\dagger}\right) - \overline{Q}_{sem}\left(\widetilde{x}; \beta^{\dagger}, \gamma^{\dagger}, \tau^{\dagger}\right)$, the same argument yields $J_{sem}\left(b\right)$ equal to

$$
\begin{aligned}
&E\left[\left\{\int_{\mathcal{X}_{pos}} \widetilde{b}\left(x, Z\right) \widetilde{Q}_{sem}\left(\widetilde{x}; \beta^{\dagger}, \gamma^{\dagger}, \tau^{\dagger}\right) dP_X\left(x\right)\right\}\right. \\
&\left. \times \left\{\int_{\mathcal{X}_{pos}} \widetilde{b}_{sem,opt}\left(x, Z\right) \widetilde{Q}_{sem}\left(\widetilde{x}; \beta^{\dagger}, \gamma^{\dagger}, \tau^{\dagger}\right)' dP_X\left(\widetilde{x}\right)\right\}\right] \\
&= E\left[\left\{\int_{\mathcal{X}_{pos}} \widetilde{b}\left(x, Z\right) Q_{sem}\left(\widetilde{x}; \beta^{\dagger}, \gamma^{\dagger}, \tau^{\dagger}\right) dP_X\left(x\right)\right\}\right. \\
&\left. \left\{\int_{\mathcal{X}_{pos}} \widetilde{b}_{sem,opt}\left(x, Z\right) Q_{sem}\left(\widetilde{x}; \beta^{\dagger}, \gamma^{\dagger}, \tau^{\dagger}\right)' dP_X\left(\widetilde{x}\right)\right\}\right] \\
&= E\left\{Q_{sem}\left(b\right) Q_{sem}\left(b_{par,opt}\right)'\right\}.
\end{aligned}
$$

Now, with $var_A\left(\widehat{\beta}_{\cdot}\left(b\right)\right)$ denoting the asymptotic variance of $\widehat{\beta}_{\cdot}\left(b\right)$, we have that from expansion ((32) in ORR-I)

$$
var_A\left(\widehat{\beta}_{\cdot}\left(b\right)\right) = var\left\{E\left[Q_{\cdot}\left(b\right) Q_{\cdot}\left(b_{\cdot,opt}\right)'\right]^{-1} Q_{\cdot}\left(b\right)\right\}
$$

and consequently

$$\begin{aligned}
&cov_A\left(\widehat{\beta}_{\cdot}(b), \widehat{\beta}_{\cdot}(b_{\cdot,opt})\right) = \\
&= E\left[Q_{\cdot}(b) Q_{\cdot}(b_{\cdot,opt})'\right]^{-1} cov\left(Q_{\cdot}(b), Q_{\cdot}(b_{\cdot,opt})\right) E\left[Q_{\cdot}(b_{\cdot,opt})^{\otimes 2}\right]^{-1} \\
&= E\left[Q_{\cdot}(b_{\cdot,opt})^{\otimes 2}\right]^{-1} = var_A\left(\widehat{\beta}_{\cdot}(b_{\cdot,opt})\right).
\end{aligned}$$

Thus, $0 \leq var_A\left(\widehat{\beta}_{\cdot}(b) - \widehat{\beta}_{\cdot}(b_{\cdot,opt})\right) = var_A\left(\widehat{\beta}_{\cdot}(b)\right) + var_A\left(\widehat{\beta}_{\cdot}(b_{\cdot,opt})\right) - 2cov_A\left(\widehat{\beta}_{\cdot}(b), \widehat{\beta}_{\cdot}(b_{\cdot,opt})\right) = var_A\left(\widehat{\beta}_{\cdot}(b)\right) - var_A\left(\widehat{\beta}_{\cdot}(b_{\cdot,opt})\right)$ which concludes the proof.

3 Confidence Set for $x_{opt}(z)$ when $\mathcal{X}$ is Finite and $h_{\cdot}(z, x; \beta)$ is Linear in β

We first prove the assertion that the computation of the confidence set B_b entails an algorithm for determining if the intersection of $\#(\mathcal{X}) - 1$ half spaces in $\mathbb{R}^p$ and a ball in $\mathbb{R}^p$ centered at the origin is non-empty. To do so, first note that linearity implies that $h_{\cdot}(z, x; \beta) = \sum_{j=1}^{p} s_j(x, z)\beta_j$ for some fixed functions $s_j, j = 1, ..., p$. Let $N = \#(\mathcal{X})$ and write $\mathcal{X} = \{x_1, ..., x_N\}$. The point x_l is in B_b iff

$$\text{there exists } \beta \text{ in } C_b: \ \sum_{j=1}^{p}\left[s_j(x_l, z) - s_j(x_k, z)\right]\beta_j \geq 0 \text{ for all } x_k \in \mathcal{X} - \{x_l\}. \tag{6}$$

Define the $p \times 1$ vector $\mathbf{v}_l^k$ whose j^{th} entry is equal to $s_j(x_l, z) - s_j(x_k, z)$, $j = 1, ..., p$. Define also the vectors $\mathbf{v}_l^{*k} = \mathbf{v}_l^{k\prime}\widehat{\Gamma}_{\cdot}(b)$ and the constants $a_l^k = \mathbf{v}_l^{k\prime}\widehat{\beta}_{\cdot}\left(b, \widehat{d}^b_{\cdot,opt}\right)$. Then $\sum_{j=1}^{p}\left[s_j(x_l, z) - s_j(x_k, z)\right]\beta_j > 0$ iff $\mathbf{v}_l^{*k}\widehat{\Gamma}_{\cdot}(b)^{-1/2} \times \left(\beta - \widehat{\beta}_{\cdot}\left(b, \widehat{d}^b_{\cdot,opt}\right)\right) > a_l^k$. Noting that β in C_b iff $\widehat{\Gamma}_{\cdot}(b)^{-1/2}\left(\beta - \widehat{\beta}_{\cdot}\left(b, \widehat{d}^b_{\cdot,opt}\right)\right)$ is in the ball

$$\mathcal{U} \equiv \left\{\mathbf{u} \in \mathbb{R}^p : \mathbf{u}'\mathbf{u} \leq \chi^2_{p,1-\alpha}\right\}$$

we conclude that the condition in the display (6) is equivalent to

$$\text{there exists } \mathbf{u} \text{ in } \mathcal{U} \text{ such that } \mathbf{v}_l^{*k\prime}\mathbf{u} > a_l^k \text{ for } k = 1, ..., N, k \neq l.$$

The set $\left\{\mathbf{u} \in \mathbb{R}^p : \mathbf{v}_l^{*k\prime}\mathbf{u} = a_l^k\right\}$ is a hyper-plane in $\mathbb{R}^p$ which divides the Euclidean space $\mathbb{R}^p$ into two half-spaces, one of which is $\left\{\mathbf{u} \in \mathbb{R}^p : \mathbf{v}_l^{*k\prime}\mathbf{u} > a_l^k\right\}$. Thus, the condition in the last display imposes that the intersection of $N-1$ half-spaces (each one defined by the condition $\mathbf{v}_l^{*k\prime}\mathbf{u} > a_l^k$ for each k) and the ball $\mathcal{U}$ is non-empty.

Turn now to the construction of a confidence set B_b^* that includes B_b. Our construction relies on the following Lemma.

Lemma. Let

$$\mathcal{D} = \left\{\mathbf{u} \in \mathbf{R}^p : (\mathbf{u}-\mathbf{u}_0)^\prime \Sigma^{-1} (\mathbf{u}-\mathbf{u}_0) \leq c_0\right\}$$

where $\mathbf{u}_0$ is a fixed $p\times 1$ real valued vector and Σ is a fixed non-singular $p\times p$ matrix.

Let $\boldsymbol{\alpha}$ be a fixed, non-null, $p\times 1$ real valued vector. Let $\tau_0 \equiv \boldsymbol{\alpha}^\prime \mathbf{u}_0$ and $\boldsymbol{\alpha}^* = \Sigma^{1/2}\boldsymbol{\alpha}$. Assume that $\alpha_1 \neq 0$. Let, $\mathbf{v}_1^*$ be the $p\times 1$ vector $\left(-\alpha_1^{*-1}\tau_0, 0, ..., 0\right)^\prime$. Let Υ be the linear space generated by the $p\times 1$ vectors $\mathbf{v}_2^* = \left(\alpha_1^{*-1}\alpha_2^*, 1, 0, 0, ..., 0\right)^\prime$, $\mathbf{v}_3^* = \left(\alpha_1^{*-1}\alpha_3^*, 0, 1, 0, .., 0\right)^\prime$, ..., $\mathbf{v}_p^* = \left(\alpha_1^{*-1}\alpha_p^*, 0, 0, 0, ..., 1\right)^\prime$ and define

$$\begin{aligned}\mathbf{v}_{1,proj}^* &= \mathbf{v}_1^* - \Pi\left[\mathbf{v}_1^*|\Upsilon\right] \\ &= \mathbf{v}_1^* - \mathbf{V}^*\left(\mathbf{V}^{*\prime}\mathbf{V}^*\right)^{-1}\mathbf{V}^{*\prime}\mathbf{v}_1^*\end{aligned}$$

where

$$\mathbf{V}^* = \left(\mathbf{v}_2^*, ..., \mathbf{v}_p^*\right).$$

Then there exists $\mathbf{u} \in \mathcal{D}$ satisfying

$$\boldsymbol{\alpha}^\prime \mathbf{u} = \mathbf{0}$$

if and only if

$$c_0 - \left\|\mathbf{v}_{1,proj}^*\right\|^2 \geq 0.$$

Proof

$$\boldsymbol{\alpha}^\prime \mathbf{u} = \mathbf{0} \;\Leftrightarrow \boldsymbol{\alpha}^\prime \Sigma^{1/2}\Sigma^{-1/2}\left(\mathbf{u}-\mathbf{u}_0\right) = -\boldsymbol{\alpha}^\prime \mathbf{u}_0.$$

Then, with $\tau_0 \equiv -\boldsymbol{\alpha}^\prime \mathbf{u}_0$ and $\boldsymbol{\alpha}^* = \Sigma^{1/2}\boldsymbol{\alpha}$, we conclude that there exists $\mathbf{u} \in \mathcal{D}$ satisfying $\boldsymbol{\alpha}^\prime \mathbf{u} = \mathbf{0}$ if and only if there exists $\mathbf{u}^* \in \mathbf{R}^p$ such that

$$\mathbf{u}^{*\prime}\mathbf{u}^* \leq c_0 \text{ and } -\boldsymbol{\alpha}^{*\prime}\mathbf{u}^* = \tau_0.$$

Now, by the assumption $\alpha_1^* \neq 0$ we have $-\boldsymbol{\alpha}^{*\prime}\mathbf{u}^* = \tau_0$ iff $u_1 = -\alpha_1^{*-1} \times \left[\tau_0 + \sum_{j=2}^p \alpha_j^* u_j^*\right]$. Thus, the collection of all vectors $\mathbf{u}^*$ satisfying $-\boldsymbol{\alpha}^{*\prime}\mathbf{u}^* = \tau_0$ is the linear variety

$$\mathbf{v}_1^* + \Upsilon = \mathbf{v}_{1,proj}^* + \Upsilon$$

DOI: 10.2202/1557-4679.1242

where $\mathbf{v}_j^{*\prime}s$ and Υ are defined in the statement of the lemma. The vector $\mathbf{v}_{1,proj}^*$ is the residual from the (Euclidean) projection of $\mathbf{v}_1^*$ into the space Υ.

Thus, $-\boldsymbol{\alpha}^{*\prime}\mathbf{u}^* = \tau_0$ iff $\mathbf{u}^* = \mathbf{v}_{1,proj}^* + \mathbf{v}_\Upsilon^*$ for some $\mathbf{v}_\Upsilon^* \in \Upsilon$. Consequently, by the orthogonality of $\mathbf{v}_{1,proj}^*$ with Υ we have that for $\mathbf{u}^*$ satisfying $-\boldsymbol{\alpha}^{*\prime}\mathbf{u}^* = \tau_0$ it holds that

$$\begin{aligned}\mathbf{u}^{*\prime}\mathbf{u}^* &= \|\mathbf{u}^*\|^2 \\ &= \left\|\mathbf{v}_{1,proj}^*\right\|^2 + \|\mathbf{v}_\Upsilon^*\|^2 .\end{aligned}$$

Therefore, since $\|\mathbf{v}_\Upsilon^*\|^2$ is unrestricted,

$$\mathbf{u}^{*\prime}\mathbf{u}^* \leq c_0 \text{ for some } \mathbf{u}^* \text{ satisfying } -\boldsymbol{\alpha}^{*\prime}\mathbf{u}^* = \tau_0$$

if and only if

$$c_0 - \left\|\mathbf{v}_{1,proj}^*\right\|^2 \geq 0. \tag{7}$$

This concludes the proof of the Lemma.

To construct the set B_b^* we note that the condition in the display (6) implies the negation, for every subset $\mathcal{X}_{(-l)}$ of $\mathcal{X} - \{x_l\}$, of the statement

$$\sum_{j=1}^{p} \sum_{k \in \mathcal{X}_{(-l)}} \left[s_j(x_l, z) - s_j(x_k, z)\right] \beta_j < 0 \text{ for all } \beta \in C_b. \tag{8}$$

Thus, suppose that for a given x_l we find that (8) holds for some subset $\mathcal{X}_{(-l)}$ of $\mathcal{X} - \{x_l\}$, then we know that x_l cannot be in B_b. The proposed confidence set B_b^* is comprised by the points in $\mathcal{X}$ for which condition (8) cannot be negated for all subsets $\mathcal{X}_{(-l)}$. The set B_b^* is conservative (i.e. it includes B_b but is not necessarily equal to B_b) because the simultaneous negation of the statement (8) for all $\mathcal{X}_{(-l)}$ does not imply the statement (6). To check if condition (8) holds for any given subset $\mathcal{X}_{(-l)}$ and x_l, we apply the result of Lemma as follows. We define the vector $\boldsymbol{\alpha} \in \mathbb{R}^p$ whose j^{th} component is equal to $\sum_{k \in \mathcal{X}_{(-l)}} \left[s_j(x_l, z) - s_j(x_k, z)\right], j = 1, ..., p$ and the vector $\mathbf{u}_0 = \widehat{\beta}_\cdot\left(b, \widehat{d}_{\cdot,opt}^b\right) \in \mathbb{R}^p$. We also define the constant $c_0 = \chi_{p,1-\alpha}^2$, and the matrix $\Sigma = \widehat{\Gamma}_\cdot(b)$. We compute the vectors $\boldsymbol{\alpha}^* = \Sigma^{1/2}\boldsymbol{\alpha}$, $\mathbf{v}_1^*, ..., \mathbf{v}_p^*$ and the matrix $\mathbf{V}^*$ as defined in Lemma. We then check if the condition (7) holds. If it holds then this implies that the hyperplane comprised by the set of β's that satisfy the condition in display (8) with the $<$ sign replaced by the $=$ sign, intersects the confidence ellipsoid C_b, in which case we know that (8) is false. If it does not hold, then we check if condition

$$\sum_{j=1}^{p} \sum_{k \in \mathcal{X}_{(-l)}} \left[s_j(x_l, z) - s_j(x_k, z)\right] \widehat{\beta}_\cdot\left(b, \widehat{d}_{\cdot,opt}^b\right)_j < 0 \tag{9}$$

holds. If (9) does not hold, then we conclude that (8) is false for this choice of $\mathcal{X}_{(-l)}$. If (9) holds, then we conclude that (8) is true and we then exclude x_l from the set B_b^*.

4 Positivity Assumption: Example

Suppose that $K = 1$ and that $R_k = R_k^g = 1$ with probability 1 for $k = 0, 1$, so that no subject dies in neither the actual world nor in the hypothetical world in which g is enforced in the population. Thus, for $k = 0, 1$, $O_k = L_k$ since both T_k and R_k are deterministic and hence can be ignored. Suppose that L_k and A_k are binary variables (and so are therefore A_k^g and L_k^g) and that the treatment regime g specifies that

$$g_0(l_0) = 1 - l_0 \text{ and } g_1(l_0, l_1) = l_0(1 - l_1).$$

Assume that

$$0 < \mathbb{P}(L_0^g = l_0, L_1^g = l_1) < 1, l_0 = 0, 1; l_1 = 0, 1. \tag{10}$$

Assumption PO imposes two requirements,

$$\mathbb{P}[\lambda_0(A_0^g | L_0^g) > 0] = 1 \text{ and} \tag{11}$$

$$\mathbb{P}[\lambda_1(A_1^g | L_0^g, L_1^g, A_0^g) > 0] = 1. \tag{12}$$

Because by definition of regime g, $A_0^g = 1 - L_0^g$, then requirement (11) can be re-expressed as

$$1 = \mathbb{P}(L_0^g = 0) I_{(0,1]}(\lambda_0(1|0)) + \mathbb{P}(L_0^g = 1) I_{(0,1]}(\lambda_0(0|1)).$$

Since indicators can only take the values 0 or 1 and $\mathbb{P}(L_0^g = l_0) < 1, l_0 = 0, 1$ (by assumption (10)), the preceding equality is equivalent to

$$I_{(0,1]}(\lambda_0(1|0)) = 1 \text{ and } I_{(0,1]}(\lambda_0(0|1)) = 1,$$

that is to say,

$$\lambda_0(1|0) > 0 \text{ and } \lambda_0(0|1) > 0.$$

By the definition of $\lambda_0(\cdot|\cdot)$ (see (3) in ORR-I), the last display is equivalent to

$$\mathbb{P}(A_0 = 1 | L_0 = 0) > 0 \text{ and } \mathbb{P}(A_0 = 0 | L_0 = 1) > 0. \tag{13}$$

Likewise, because $A_1^g = L_0^g (1 - L_1^g)$, and because $\mathbb{P}(L_0^g = l_0, L_1^g = l_1, A_0^g = l_0) = 0$ by the fact that $A_0^g = 1 - L_0$, requirement (12) can be re-expressed as

$$\begin{aligned} 1 &= \mathbb{P}\left(L_0^g = 0, L_1^g = 0, A_0^g = 1\right) I_{(0,1]}\left(\lambda_1\left(0|0,0,1\right)\right) \\ &+ \mathbb{P}\left(L_0^g = 0, L_1^g = 1, A_0^g = 1\right) I_{(0,1]}\left(\lambda_1\left(0|0,1,1\right)\right) \\ &+ \mathbb{P}\left(L_0^g = 1, L_1^g = 0, A_0^g = 0\right) I_{(0,1]}\left(\lambda_1\left(1|1,0,0\right)\right) \\ &+ \mathbb{P}\left(L_0^g = 1, L_1^g = 1, A_0^g = 0\right) I_{(0,1]}\left(\lambda_1\left(0|1,1,0\right)\right) \end{aligned}$$

or equivalently, (again because the events $(L_0^g = l_0, L_1^g = l_1, A_0^g = 1 - l_0)$ and $(L_0^g = l_0, L_1^g = l_1)$ have the same probability by $\mathbb{P}\left(L_0^g = l_0, L_1^g = l_1, A_0^g = l_0\right) = 0$),

$$\begin{aligned} 1 &= \mathbb{P}\left(L_0^g = 0, L_1^g = 0\right) I_{(0,1]}\left(\lambda_1\left(0|0,0,1\right)\right) + \mathbb{P}\left(L_0^g = 0, L_1^g = 1\right) \\ &\times I_{(0,1]}\left(\lambda_1\left(0|0,1,1\right)\right) + \mathbb{P}\left(L_0^g = 1, L_1^g = 0\right) I_{(0,1]}\left(\lambda_1\left(1|1,0,0\right)\right) \\ &+ \mathbb{P}\left(L_0^g = 1, L_1^g = 1\right) I_{(0,1]}\left(\lambda_1\left(0|1,1,0\right)\right). \end{aligned}$$

Under the assumption (10), the last display is equivalent to

$$\begin{aligned} &\lambda_1\left(0|0,0,1\right) > 0, \lambda_1\left(0|0,1,1\right) > 0, \\ &\lambda_1\left(1|1,0,0\right) > 0 \text{ and } \lambda_1\left(0|1,1,0\right) > 0 \end{aligned}$$

which, by the definition of $\lambda_0\left(\cdot|\cdot,\cdot,\cdot\right)$ in ((3), ORR-I), is, in turn, the same as

$$\begin{aligned} &\mathbb{P}\left(A_1 = 0|L_0 = 0, L_1 = 0, A_0 = 1\right) > 0,\ \mathbb{P}\left(A_1 = 0|L_0 = 0, L_1 = 1, A_0 = 1\right) > 0 \\ &\mathbb{P}\left(A_1 = 1|L_0 = 1, L_1 = 0, A_0 = 0\right) > 0,\ \mathbb{P}\left(A_1 = 0|L_0 = 1, L_1 = 1, A_0 = 0\right) > 0\,. \end{aligned} \tag{14}$$

We conclude that in this example, the assumption PO is equivalent to the conditions (13) and (14). We will now analyze what these conditions encode.

Condition (13) encodes two requirements:

i) the requirement that in the actual world there exist subjects with $L_0 = 1$ and $L_0 = 0$ (i.e. that the conditioning events $L_0 = 1$ and $L_0 = 0$ have positive probabilities), for otherwise at least one of the conditional probabilities in (13) would not be defined, and

ii) the requirement that in the actual world there be subjects with $L_0 = 0$ that take treatment $A_0 = 1$ and subjects with $L_0 = 1$ that take treatment $A_0 = 0$, for otherwise at least one of the conditional probabilities in (13) would be 0.

Condition i) is automatically satisfied, i.e. it does not impose a restriction on the law of L_0, by the fact that $L_0^g = L_0$ (since baseline covariates cannot be affected by interventions taking place after baseline) and the fact that we have assumed that $\mathbb{P}(L_0^g = l_0) > 0, l_0 = 0, 1$.

Condition ii) is indeed a non-trivial requirement and coincides with the interpretation of the PO assumption given in section 3.1 for the case $k = 0$. Specifically, in the world in which g were to be implemented there would exist subjects with $L_0 = 0$. In such world the subjects with $L_0 = 0$ would take treatment $A_0^g = 1$, then the PO assumption for $k = 0$ requires that in the actual world there also be subjects with $L_0 = 0$ that at time 0 take treatment $A_0 = 1$. Likewise the PO condition also requires that for $k = 0$ the same be true with 0 and 1 reversed in the right hand side of each of the equalities of the preceding sentence. A key point is that (11) does not require that in the observational world there be subjects with $L_0 = 0$ that take $A_0 = 0$, nor subjects with $L_0 = 1$ that take $A_1 = 1$. The intuition is clear. If we want to learn from data collected in the actual (observational) world what would happen in the hypothetical world in which everybody obeyed regime g, we must observe people in the study that obeyed the treatment at every level of L_0 for otherwise if, say, nobody in the actual world with $L_0 = 0$ obeyed regime g there would be no way to learn what the distribution of the outcomes for subjects in that stratum would be if g were enforced. However, we don't care that there be subjects with $L_0 = 0$ that do not obey g, i.e. that take $A_0 = 0$, because data from those subjects are not informative about the distribution of outcomes when g is enforced.

Condition (14) encodes two requirements:

iii) the requirement that in the actual world there be subjects in the four strata $(L_0 = 0, L_1 = 0, A_0 = 1), (L_0 = 0, L_1 = 1, A_0 = 1), (L_0 = 1, L_1 = 0, A_0 = 0)$ and $(L_0 = 1, L_1 = 1, A_0 = 0)$ (i.e. that the conditioning events in the display (14) have positive probabilities), for otherwise at least one of the conditional probabilities would not be defined, and

iv) the requirement that in the actual world there be subjects in every one of the strata $(L_0 = 0, L_1 = 0, A_0 = 1)$, $(L_0 = 0, L_1 = 1, A_0 = 1)$, $(L_0 = 1, L_1 = 1, A_0 = 0)$ that have $A_1 = 0$ at time 1 and the requirement that there be subjects in stratum $(L_0 = 1, L_1 = 0, A_0 = 0)$ that have $A_1 = 1$ at time 1, for otherwise at least one of the conditional probabilities in (14) would be 0.

Given condition ii) and the sequential randomization (SR) and consistency (C) assumptions, condition iii) is automatically satisfied, i.e. it does not im-

DOI: 10.2202/1557-4679.1242

pose a further restriction on the joint distribution of (L_0, L_1, A_0). To see this, first note that by condition (ii) the strata $(L_0 = 0, A_0 = 1)$ and $(L_0 = 1, A_0 = 0)$ are non-empty. So condition (iii) is satisfied provided

$$\mathbb{P}(L_1 = l_1 | L_0 = 0, A_0 = 1) > 0 \text{ and } \mathbb{P}(L_1 = l_1 | L_0 = 1, A_0 = 0) > 0 \text{ for } l_1 = 0, 1.$$

But

$$\begin{aligned}
\mathbb{P}(L_1 = l_1 | L_0 = 0, A_0 = 1) &= \mathbb{P}(L_1^g = l_1 | L_0 = 0, A_0 = 1) \text{ by assumption (C)} \\
&= \mathbb{P}(L_1^g = l_1 | L_0 = 0) \text{ by assumption (SR)} \\
&= \mathbb{P}(L_1^g = l_1 | L_0^g = 0) \text{ by assumption (C)}
\end{aligned}$$

and $\mathbb{P}(L_1^g = l_1 | L_0^g = 0) > 0$ by (10). An analogous argument shows that $\mathbb{P}(L_1 = l_1 | L_0 = 1, A_0 = 0) > 0$. Finally, condition (iv) is a formalization our interpretation of assumption PO in section 3.1 for $k = 1$. In the world in which g was implemented there would exist subjects that would have all four combination of values for (L_0^g, L_1^g). However, subjects with $L_0^g = l_0$ will only have $A_0^g = 1 - l_0$, so in this hypothetical world we will see at time 1 only four possible recorded histories, $(L_0^g = 0, L_1^g = 0, A_0^g = 1)$, $(L_0^g = 0, L_1^g = 1, A_0^g = 1)$, $(L_0^g = 1, L_1^g = 0, A_0^g = 0)$ and $(L_0^g = 1, L_1^g = 1, A_0^g = 0)$. In this hypothetical world subjects with any of the first three possible recorded histories will take $A_1^g = 0$ and subjects with the last one will take $A_1^g = 1$. Thus, in the actual world we must require that there be subjects in each of the first three strata $(L_0 = 0, L_1 = 0, A_0 = 1)$, $(L_0 = 0, L_1 = 1, A_0 = 1)$, $(L_0 = 1, L_1 = 0, A_0 = 0)$ that take $A_1 = 0$ and subjects in the stratum $(L_0 = 1, L_1 = 1, A_0 = 0)$ that take $A_1 = 1$. A point of note is that we don't make any requirement about the existence of subjects in strata other than the four mentioned in (iii) or about the treatment that subjects in these remaining strata take. The reason is that subjects that are not in the four strata of condition (iii) have already violated regime g at time 0 so they are uninformative about the outcome distribution under regime g.

References

Orellana, L., Rotnitzky, A. and Robins J.M. (2010). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: Main content. *The International Journal of Biostatistics*, Vol. 6: Iss. 2, Article 7.