Biostatistics (2014), **0**, 0, *pp.* 1–17
doi:10.1093/biostatistics/BIOSTS˙13270˙R2˙SupplementaryNew

Characterizing Expected Benefits of Biomarkers in Treatment Selection - Supplementary Materials

Ying Huang[1*], Eric Laber[2], and Holly Janes[1]
[1]*Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N., Seattle WA*

[2]*Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, NC*

*yhuang@fhcrc.org

Supplementary Figures

In Figures 1(a)(b), we show the expected benefit of two treatment-selection markers and lower and upper bounds for perfect treatment selection derived from corresponding marker-based risk model. Marker 1 (Figure 1(a)) has small expected benefit with a large potential for improvement. For example, at a cost ratio 0.05, its expected benefit of 0.005 is far from the perfect selection rule: a perfect selection rule can have an expected benefit 8.8-26.9 times that of Marker 1; corresponding standardized expected benefit for Marker 1 at cost ratio 0.05 ranges between 3.7% and 11.4% (Figure 1(c)). In contrast, there is less potential to improve over a better marker (Marker 2) (Figure 1(b)). At a cost ratio 0.05, a perfect selection rule can have expected benefit 1.7-2.8 times that of Marker 2, which has expected benefit 0.05 and standardized value ranging from 35.7% to 58.0% (Figure 1(d)).

*To whom correspondence should be addressed.

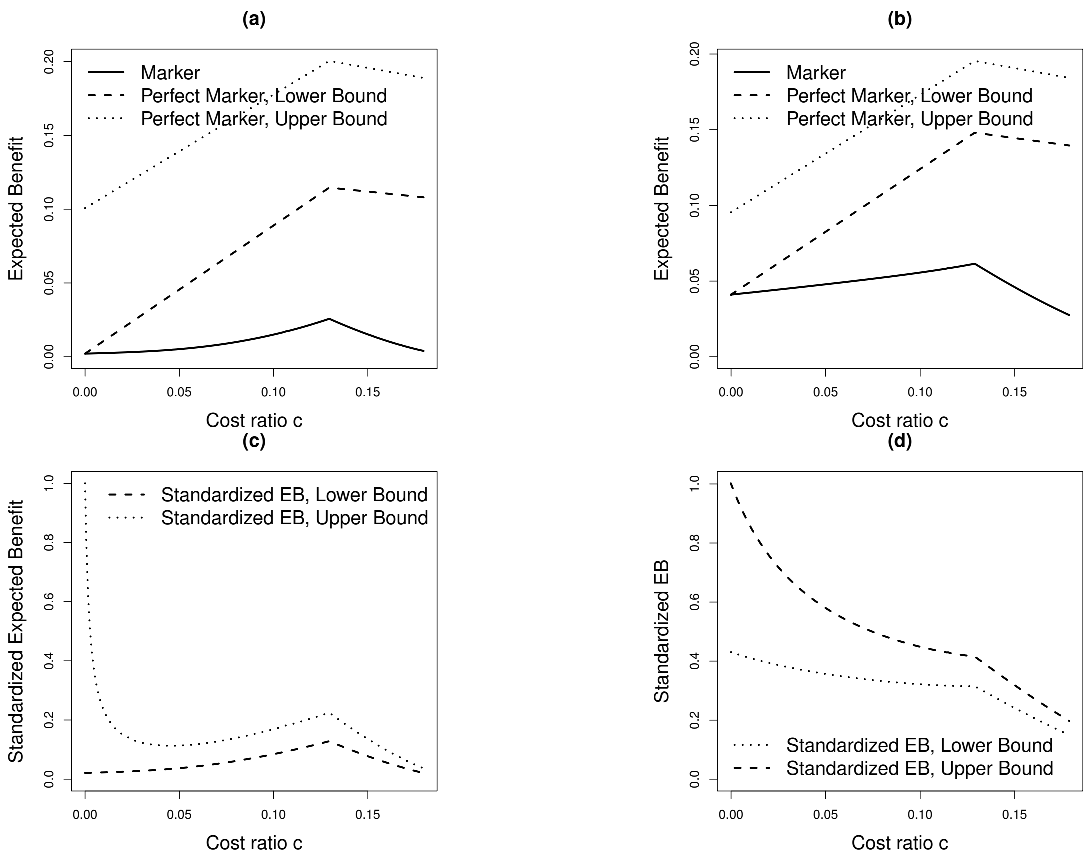

Fig. 1. Expected benefit curves for Marker 1 (a) and Marker 2 (b) and bounds for perfect treatment-selection rule, and corresponding standardized expected benefit curves for Marker 1 (c) and Marker 2 (d). Disease prevalence among untreated and treated subjects is 0.25 and 0.12 respectively. Each marker Y follows a N(2,1) distribution and its relationship to disease status is described by a logistic risk model, $\text{logit}P(D=1|Y,T)=\beta_0+\beta_1 T+\beta_2 Y+\beta_3 YT$, where T is an indicator of treatment assignment. For Marker 1, $\beta_0=0.69, \beta_1=0.2, \beta_2=-1, \beta_3=-1$. For Marker 2, $\beta_0=-0.158, \beta_1=3.495, \beta_2=-0.5, \beta_3=-4$. Note that in this setting the optimal treatment strategy in the absence of marker information is to treat everyone if $c<0.13$ and treat no one otherwise.

Supplementary Tables

Table 1: Bootstrap standard error and coverage based on (naive) robust estimators of expected benefit. NPAR1 and NPAR2 correspond to $A(Y) = I(\hat{\Delta}^\star > \delta)$ with δ estimated to maximize empirical or augmented estimate of EB; WLOGIS1, WLOGIS2 correspond to rule $A(Y) = I(\alpha_0 + \alpha_1 Y > 0)$ where α_0, α_1 were estimated based on weighted logistic regression with empirical weight or augmented weight respectively; WCLASS1, WCLASS2 correspond to rule $A(Y) = I(\alpha_0 + \alpha_1 Y > 0)$ where α_0, α_1 were estimated based on grid search to minimize weighted misclassification error with empirical weight or augmented weight respectively. Results are based on 1000 Monte-Carlo simulations and 200 bootstrap samples per simulation. Wald CI is constructed using bootstrap SE estimate. Bias correction is applied to adaptive CI where bias is estimated as the difference between naive estimate and cross-validated estimate (using 3-fold random CV).

			Cost ratio c				
Estimator	Type	N	0	0.105	0.125	0.145	0.175
			$SE \times \sqrt{N}$				
NPAR1		200	0.40	0.39	0.39	0.38	0.38
		500	0.44	0.42	0.42	0.42	0.41
		2000	0.46	0.46	0.45	0.46	0.46
NPAR2		200	0.40	0.38	0.38	0.38	0.37
		500	0.42	0.40	0.40	0.41	0.41
		2000	0.43	0.43	0.43	0.46	0.47
WLOGIS1		200	0.50	0.45	0.44	0.41	0.36
		500	0.54	0.52	0.52	0.50	0.42
		2000	0.57	0.52	0.55	0.63	0.48
WLOGIS2		200	0.47	0.44	0.43	0.41	0.37
		500	0.49	0.46	0.47	0.47	0.41
		2000	0.49	0.45	0.48	0.55	0.46
WCLASS1		200	0.44	0.40	0.39	0.38	0.36
		500	0.48	0.45	0.45	0.44	0.42
		2000	0.51	0.50	0.48	0.50	0.51
WCLASS2		200	0.45	0.40	0.39	0.38	0.36
		500	0.47	0.43	0.43	0.43	0.42
		2000	0.48	0.46	0.45	0.48	0.50
			Coverage of 95% bootstrap				
NPAR1	Ordinary	200	88.47	93.41	93.14	94.33	92.96
		500	91.30	94.43	95.44	95.57	89.82
		2000	93.13	92.06	95.73	92.08	91.48
	Adaptive	200	91.49	94.24	93.60	95.88	98.17
		500	92.21	96.19	95.34	96.06	96.50
		2000	92.61	96.03	97.50	96.56	95.22
NPAR2	Ordinary	200	90.94	94.88	93.60	94.33	92.68
		500	92.82	95.21	96.15	94.78	88.76
		2000	93.65	93.00	95.83	91.15	92.11
	Adaptive	200	92.50	94.33	93.78	95.24	97.80
		500	93.63	96.29	96.05	96.85	96.70
		2000	94.17	96.66	97.08	96.46	94.50
WLOGIS1	Ordinary	200	89.48	80.15	75.30	91.40	98.08

		500	92.01	91.60	87.23	89.26	95.58
		2000	94.38	96.87	96.46	91.88	97.20
	Adaptive	200	94.33	92.59	92.31	96.52	99.09
		500	95.25	94.24	94.53	97.83	98.60
		2000	94.17	95.19	97.29	96.25	98.03
WLOGIS2	Ordinary	200	92.50	89.39	83.53	90.12	98.17
		500	94.34	96.48	91.89	89.95	96.06
		2000	94.59	95.40	94.79	92.71	93.35
	Adaptive	200	93.41	93.50	93.96	96.16	97.99
		500	93.53	96.19	95.95	97.54	98.20
		2000	94.48	97.39	97.50	96.46	97.61
WCLASS1	Ordinary	200	89.39	89.75	87.10	93.05	95.06
		500	91.41	93.26	92.81	93.79	91.83
		2000	93.76	93.00	94.90	94.06	92.42
	Adaptive	200	90.39	91.22	89.57	90.67	95.88
		500	90.50	94.73	94.43	94.38	91.10
		2000	92.40	95.61	97.40	96.35	92.52
WCLASS2	Ordinary	200	91.67	92.77	89.84	91.67	94.60
		500	93.63	95.61	93.82	94.58	90.97
		2000	93.96	93.73	95.42	93.96	92.00
	Adaptive	200	93.05	93.41	92.22	93.05	93.87
		500	92.82	96.68	95.95	95.96	92.60
		2000	93.65	95.72	97.40	96.35	91.90

Table 2: MEAN (SD) of expected benefit of a derived treatment-selection rule in the population and MEAN(SD) of naive and cross-validated estimate of corresponding expected benefit when risk model is misspecified. Results are based on 5,000 Monte-Carlo Simulations.

N	TYPE	PAR*	NPAR1*	NPAR2*	WLOGIS1*	WLOGIS2*	WCLASS1*	WCLASS2*
				cost ratio c=0				
200	True	0.0824 (0.0236)	0.0985 (0.0122)	0.0989 (0.0112)	0.069 (0.0394)	0.083 (0.025)	0.0973 (0.0152)	0.0993 (0.01)
	Naive	0.0635 (0.0326)	0.109 (0.0304)	0.1091 (0.032)	0.0713 (0.0482)	0.0787 (0.0351)	0.1085 (0.0349)	0.1082 (0.0294)
	CV	0.0648 (0.0382)	0.0814 (0.0358)	0.0827 (0.0368)	0.0478 (0.0491)	0.0642 (0.0377)	0.0795 (0.0423)	0.0834 (0.0339)
500	True	0.0867 (0.0136)	0.1025 (0.0048)	0.1025 (0.0048)	0.0846 (0.0247)	0.0899 (0.014)	0.102 (0.0057)	0.1025 (0.0047)
	Naive	0.0661 (0.0229)	0.1092 (0.0203)	0.1093 (0.0212)	0.0839 (0.0309)	0.087 (0.0202)	0.1091 (0.0233)	0.1088 (0.0197)
	CV	0.0797 (0.0216)	0.0955 (0.022)	0.0955 (0.0229)	0.0725 (0.0319)	0.0819 (0.0212)	0.0947 (0.0255)	0.0958 (0.0214)
2000	True	0.0875 (0.0068)	0.1045 (0.0019)	0.1045 (0.0019)	0.0938 (0.0106)	0.0919 (0.0069)	0.1044 (0.0022)	0.1045 (0.0019)
				cost ratio c=0.0069				
	Naive	0.0696 (0.0128)	0.1087 (0.011)	0.1088 (0.0114)	0.0939 (0.0113)	0.0916 (0.0091)	0.1088 (0.0126)	0.1083 (0.0108)
	CV	0.0865 (0.0102)	0.1032 (0.0116)	0.1032 (0.012)	0.0916 (0.0116)	0.0906 (0.0091)	0.103 (0.0132)	0.1032 (0.0113)
200	True	0.084 (0.0236)	0.1004 (0.0125)	0.1009 (0.0116)	0.0707 (0.0391)	0.0843 (0.0252)	0.0992 (0.0145)	0.1012 (0.0103)
	Naive	0.0644 (0.0324)	0.1086 (0.0305)	0.1087 (0.032)	0.0702 (0.0482)	0.0774 (0.0355)	0.1083 (0.0348)	0.1078 (0.0294)
	CV	0.0643 (0.0385)	0.0809 (0.0361)	0.0823 (0.0369)	0.0471 (0.0494)	0.0633 (0.0381)	0.0791 (0.0415)	0.083 (0.0339)
500	True	0.088 (0.0136)	0.1044 (0.0049)	0.1045 (0.0048)	0.0851 (0.025)	0.0907 (0.0143)	0.104 (0.0053)	0.1045 (0.0048)
	Naive	0.0677 (0.0225)	0.1095 (0.0202)	0.1095 (0.0211)	0.0822 (0.0315)	0.0859 (0.021)	0.1094 (0.023)	0.1091 (0.0195)
	CV	0.0793 (0.0222)	0.0957 (0.022)	0.0957 (0.0228)	0.0716 (0.0323)	0.0812 (0.0218)	0.0949 (0.0251)	0.096 (0.0213)
2000	True	0.0887 (0.0068)	0.1066 (0.0018)	0.1066 (0.0018)	0.093 (0.0115)	0.0925 (0.0071)	0.1064 (0.0021)	0.1066 (0.0018)
				cost ratio c=0.0269				
	Naive	0.0723 (0.0124)	0.11 (0.0107)	0.1101 (0.0112)	0.0923 (0.0124)	0.0914 (0.0095)	0.1101 (0.0123)	0.1096 (0.0105)
	CV	0.0869 (0.0104)	0.1045 (0.0113)	0.1044 (0.0117)	0.0902 (0.0124)	0.0905 (0.0094)	0.1043 (0.0129)	0.1044 (0.011)
200	True	0.0892 (0.0229)	0.106 (0.0133)	0.1067 (0.012)	0.0753 (0.0378)	0.0884 (0.0251)	0.1054 (0.0151)	0.1071 (0.0105)
	Naive	0.0658 (0.0318)	0.1063 (0.0313)	0.1063 (0.0327)	0.0647 (0.0487)	0.0724 (0.0372)	0.1061 (0.0351)	0.1054 (0.03)
	CV	0.0615 (0.0396)	0.0776 (0.0374)	0.0792 (0.038)	0.0442 (0.0496)	0.0596 (0.0392)	0.0768 (0.0421)	0.08 (0.0348)
500	True	0.0923 (0.0133)	0.1104 (0.005)	0.1105 (0.0048)	0.0858 (0.0253)	0.0935 (0.0148)	0.1102 (0.0053)	0.1105 (0.0048)
	Naive	0.0704 (0.0216)	0.1079 (0.0207)	0.108 (0.0216)	0.0743 (0.0343)	0.0808 (0.0242)	0.1079 (0.0232)	0.1075 (0.0199)
	CV	0.0761 (0.0245)	0.0938 (0.0226)	0.094 (0.0233)	0.0661 (0.0341)	0.0767 (0.0244)	0.0934 (0.0251)	0.0942 (0.0217)
2000	True	0.0928 (0.0066)	0.1126 (0.0018)	0.1126 (0.0018)	0.0903 (0.013)	0.0947 (0.0073)	0.1125 (0.0021)	0.1126 (0.0018)
				cost ratio c=0.0469				
	Naive	0.0765 (0.0111)	0.1099 (0.0105)	0.11 (0.011)	0.0834 (0.0172)	0.0875 (0.0121)	0.11 (0.0118)	0.1095 (0.0101)
	CV	0.0849 (0.0123)	0.1043 (0.011)	0.1043 (0.0115)	0.0819 (0.017)	0.0867 (0.012)	0.1042 (0.0123)	0.1043 (0.0106)
200	True	0.0753 (0.0216)	0.0918 (0.014)	0.0928 (0.0126)	0.0606 (0.0352)	0.0733 (0.0242)	0.0916 (0.0154)	0.0931 (0.011)

	Naive	0.0655 (0.0311)	0.1019 (0.0326)	0.1018 (0.0341)	0.0573 (0.0491)	0.0657 (0.0392)	0.1017 (0.0361)	0.1009 (0.0313)
	CV	0.057 (0.0407)	0.0721 (0.0392)	0.074 (0.0396)	0.0395 (0.0495)	0.0541 (0.0403)	0.0715 (0.0435)	0.0747 (0.0364)
500	True	0.0776 (0.0128)	0.0966 (0.0054)	0.0968 (0.0051)	0.067 (0.0244)	0.0771 (0.0147)	0.0964 (0.0059)	0.0968 (0.005)
	Naive	0.0696 (0.0211)	0.1026 (0.0224)	0.1026 (0.0233)	0.0629 (0.0372)	0.0723 (0.0277)	0.1026 (0.0244)	0.1022 (0.0215)
	CV	0.0698 (0.0272)	0.0882 (0.0243)	0.0884 (0.0249)	0.0571 (0.0363)	0.0688 (0.0275)	0.0879 (0.0262)	0.0886 (0.0232)
2000	True	0.078 (0.0064)	0.099 (0.0018)	0.099 (0.0018)	0.0693 (0.0128)	0.078 (0.0073)	0.0988 (0.0021)	0.099 (0.0018)
	Naive	0.0738 (0.0113)	0.1026 (0.0123)	0.1027 (0.0128)	0.0686 (0.0213)	0.0771 (0.0155)	0.1027 (0.0131)	0.1023 (0.0118)
	CV	0.0765 (0.015)	0.097 (0.0128)	0.097 (0.0132)	0.0675 (0.021)	0.0763 (0.0154)	0.0968 (0.0135)	0.0969 (0.0123)
					cost ratio c=0.0769			
200	True	0.0564 (0.0194)	0.0712 (0.0148)	0.0727 (0.0128)	0.041 (0.0293)	0.0527 (0.022)	0.0708 (0.0172)	0.0729 (0.012)
	Naive	0.0617 (0.0303)	0.0916 (0.0349)	0.0913 (0.0364)	0.0451 (0.0481)	0.0535 (0.0409)	0.0912 (0.0379)	0.0903 (0.0336)
	CV	0.0476 (0.0416)	0.06 (0.042)	0.062 (0.0424)	0.0309 (0.0479)	0.0435 (0.0409)	0.0592 (0.0458)	0.0625 (0.0394)
500	True	0.058 (0.0116)	0.0765 (0.006)	0.0768 (0.0051)	0.0429 (0.0209)	0.0551 (0.0138)	0.0765 (0.006)	0.0768 (0.0051)
	Naive	0.0632 (0.0209)	0.0888 (0.0249)	0.0889 (0.0256)	0.0443 (0.0381)	0.0561 (0.0302)	0.0887 (0.0265)	0.0884 (0.0238)
	CV	0.056 (0.0293)	0.0735 (0.0271)	0.074 (0.0273)	0.0404 (0.0367)	0.0528 (0.0299)	0.0734 (0.0286)	0.0742 (0.0256)
2000	True	0.0584 (0.0058)	0.0792 (0.0019)	0.0792 (0.0019)	0.0435 (0.0113)	0.0557 (0.0069)	0.079 (0.0022)	0.0792 (0.0019)
	Naive	0.0633 (0.0112)	0.0844 (0.0135)	0.0845 (0.0139)	0.0442 (0.0214)	0.0563 (0.0163)	0.0844 (0.0141)	0.084 (0.013)
	CV	0.0583 (0.0158)	0.0786 (0.0139)	0.0786 (0.0143)	0.0433 (0.0212)	0.0555 (0.0162)	0.0785 (0.0145)	0.0785 (0.0134)

PAR⋆: corresponds to the rule $A(Y) = I(\hat{\Delta} > c)$ where $\hat{\Delta}$ is the estimated risk difference based on GLM risk model
NPAR1⋆,NPAR2⋆: correspond to $A(Y) = I(\hat{\Delta} > \delta)$ with δ chosen to maximize the empirical or augmented estimate of EB;
WLOGIS1⋆, WLOGIS2⋆: correspond to rule $A(Y) = I(\alpha_0 + \alpha_1 Y > 0)$ where α_0, α_1 are estimated based on converting the problem to a weighted classification problem, which is solved using weighted logistic regression with empirical weight or augmented weight respectively;
WCLASS1⋆, WCLASS2⋆: correspond to rule $A(Y) = I(\alpha_0 + \alpha_1 Y > 0)$ where α_0, α_1 are estimated based on converting the problem to a weighted classification problem, which is solved using a grid search;
True⋆: indicates population performance of a treatment-selection rule derived from a training data.
[a] Navie for PAR: naive parametric estimate of EB for model-based treatment-selection rule rule as shown in Section 3.1

[b] CV for PAR: cross-validated nonparametric estimate of EB (as shown in equation (3.12)) for model-based treatment-selection rule.

Note: In this setting 90% data points were generated from $Y \sim N(2,1)$ with $\text{logit}P(D=1|Y) = -1.5+4.5T-3YT+0.2Y^2T-0.1Y^3$;

10% outliers were generated from $Y \sim N(4,1.5)$ with $\text{logit}P(D=1|Y) = -2+1.5T+0.2Y-3YT+0.1Y^2+0.2Y^2T+0.2Y^3-0.1Y^3$.

Disease prevalences are 0.256 and 0.229 among treated and untreated, respectively.

Table 3. Cross-Validated estimates of $EB(c)$ in DCCT example.

Cost ratio c	0	0.05	0.10	0.12
PAR(Naive)	0.005	0.019	0.035	0.028
PAR(CV)	0.0047	0.017	0.021	0.013
NPAR(CV)	0.0065	0.019	0.022	0.012
AUG(CV)	0.0048	0.014	0.017	0.0064
WLOGIS1(CV)	-0.012	-0.006	0.004	0.0006
WLOGIS2(CV)	0.015	0.023	0.024	0.015
WCLASS1 (CV)	-0.014	-0.005	0.005	-0.002
WCLASS2 (CV)	0.015	0.024	0.026	0.016

Supplementary Appendix

Appendix A: Derivation of (2.1)

Suppose we apply a treatment-selection rule $A(Y)$ to a population. The total cost, i.e., the sum of the disease and treatment cost, is $E_A(D) + P\{A(Y) = 1\} \times c = P\{A(Y) = 0\}P\{D(0) = 1|A(Y) = 0\} + P\{A(Y) = 1\}P\{D(1) = 1|A(Y) = 1\} + P\{A(Y) = 1\} \times c$, where $D(1), D(0)$ indicating potential disease outcomes if receiving or not receiving the treatment. In a randomized trial this equals

$$
\begin{aligned}
&P\{A(Y)=0\}P\{D=1|A(Y)=T=0\} + P\{A(Y)=1\}P\{D=1|A(Y)=T=1\} + P\{A(Y)=1\} \times c \\
&= P\{D=1, A(Y)=0|T=0\} + P\{D=1, A(Y)=1|T=1\} + P\{A(Y)=1\} \times c \\
&= P(D=1|T=0) - [P\{D=1, A(Y)=1|T=0\} - P\{D=1, A(Y)=1|T=1\}] + E\{A(Y)\} \times c \\
&= P(D=1|T=0) - [E\{DA(Y)|T=0\} - E\{DA(Y)|T=1\}] + E\{A(Y)\} \times c \\
&= P(D=1|T=0) - (E\,[E\{DA(Y)|Y,T=0\}|T=0] - E\,[E\{DA(Y)|Y,T=1\}|T=1]) + E\{A(Y)\} \times c \\
&= P(D=1|T=0) - (E\,[A(Y)E\{D|Y,T=0\}|T=0] - E\,[A(Y)E\{D|Y,T=1\}|T=1]) + E\{A(Y)\} \times c \\
&= P(D=1|T=0) - [E\,\{A(Y)E(D|Y,T=0)\} - E\,\{A(Y)E(D|Y,T=1)\}] + E\{A(Y)\} \times c \\
&= P(D=1|T=0) - E\,[A(Y)\,\{E(D|Y,T=0) - E(D|Y,T=1) - c\}] \\
&= \rho_0 - E\,[A(Y)\,\{\Delta(Y) - c\}]\,,
\end{aligned}
$$

where $\Delta(Y) = P\,(D=1|T=0,Y) - P\,(D=1|T=1,Y)$ is the risk difference conditional on Y between placebo and treatment arms.

Appendix B: Augmented Estimator of EB

Following the arguments in **?**, let $D(t)$ indicate the potential disease outcome if a subject received treatment t, $t = 0, 1$. Let $D(A) = D(0)\{A(Y) = 0\} + D(1)\{A(Y) = 1\}$ represent the potential outcome that would be observed if a randomly chosen subject from the population were to be assigned treatment according to rule $A(Y)$. Let $C_A = T \times A(Y) + (1 - T) \times \{1 - A(Y)\}$ be the indicator that $D(A)$ is observed. That is $D(A)$ is observed if $C_A = 1$ or $A(Y) = T$, and $D(A)$ is missing if $C_A = 0$ or $A(Y) \neq T$. Then for estimating $E\{D(A)\}$, a double-robustness estimator following the idea of **?** is

$$
\frac{1}{N}\sum_{i=1}^{N} \frac{C_{Ai}D_i}{P(C_A=1|Y_i)} - \frac{C_{Ai} - P(C_A=1|Y_i)}{P(C_A=1|Y_i)} m(Y_i;\hat{\beta}), \tag{0.1}
$$

where $P(C_A = 1|Y) = P\{A(Y) = T|Y\} = P\{T = A(Y) = 1|Y\} + P\{T = A(Y) = 0|Y\} = P(T = 1|Y) \times A(Y) + P(T = 0|Y) \times \{1 - A(Y)\}$; $m(Y_i; \beta) = E\{D(A)|Y_i\} = \mu(Y, T = 1; \beta) \times A(Y) + \mu(Y, T = 0; \beta) \times \{1 - A(Y)\}$ where $\mu(Y, T; \beta) = E(D|Y, T)$ is a working model for the risk of D conditional on Y, T with coefficient β.

In particular, note the first component of (0.1) is an inverse probability weighted estimator of $E\{D(A)\}$ which can also be written as

$$
\begin{aligned}
& \frac{1}{N}\sum_{i=1}^{N} \frac{C_{Ai}D_i}{P(T=1|Y_i) \times A(Y_i) + P(T=0|Y_i)\{1-A(Y_i)\}} \\
= & \frac{1}{N}\sum_{i=1}^{N} \frac{[T_i A(Y_i) + (1-T_i)\{1-A(Y_i)\}]\, D_i}{P(T=1|Y_i) \times A(Y_i) + P(T=0|Y_i)\{1-A(Y_i)\}} \\
= & \frac{1}{N}\sum_{i=1}^{N} \frac{D_i T_i A(Y_i)}{P(T=1|Y_i)} + \frac{1}{N}\sum_{i=1}^{N} \frac{D_i(1-T_i)\{1-A(Y_i)\}}{P(T=0|Y_i)},
\end{aligned}
$$

which in a randomized trial essentially is an empirical estimate of $P\{D = 1, A(Y) = 1|T = 1\} + P\{D = 1, A(Y) = 0|T = 0\}$. The second component is

$$
\eta \equiv -\frac{1}{N}\sum_{i=1}^{N} \frac{T_i A(Y_i) + (1-T_i) \times \{1-A(Y_i)\} - [P(T=1|Y_i)A(Y_i) + P(T=0|Y_i) \times \{1-A(Y_i)\}]}{P(T=1|Y_i)A(Y_i) + P(T=0|Y_i)\{1-A(Y_i)\}} m(Y_i\ \hat{\beta}).
$$

Finally, expected benefit is equal to $EB_A(c) = \rho_0 - [\rho_0 - \rho_1 - c]_+ - [E\{D(A)\} + P\{A(Y) = 1\} \times c]$. One can estimate $EB_A(c)$ nonparametrically, with all the terms $\rho_0, \rho_1, E\{D(A)\}, P\{A(Y) = 1\}$ estimated empirically as in (3.8); or using the augmented version (3.9) which augment the non-parametric estimate of $EB_A(c)$ with the augmented term $-\eta$.

Appendix C: Derivation of Treatment-Selection Rules by Minimizing the Weighted Classification Error

For a treatment-selection rule $A(Y) = I(\alpha_0 + \alpha_1 Y > 0)$, solution of α_0, α_1 that maximizes (3.8) equals to

$$
\begin{aligned}
(\hat\alpha_0, \hat\alpha_1) &= \operatorname{argmin}_{\alpha_0,\alpha_1} \sum_{i=1}^{N} \frac{D_i T_i I\left(\alpha_0 + \alpha_1^T Y_i > 0\right)}{N_1} + \sum_{i=1}^{N} \frac{D_i(1-T_i) I\left(\alpha_0 + \alpha_1^T Y_i \leqslant 0\right)}{N_0} + c \times \frac{I\left(\alpha_0 + \alpha_1^T Y_i > 0\right)}{N} \\
&= \operatorname{argmin}_{\alpha_0,\alpha_1} \sum_{i=1}^{n} - \left\{D_i T_i / N_1 - D_i(1-T_i)/N_0 + c/N\right\} I\left(\alpha_0 + \alpha_1^T Y_i \leqslant 0\right) \\
&= \operatorname{argmin}_{\alpha_0,\alpha_1} W_i I(\alpha_0 + \alpha_1^T Y_i \leqslant 0) \\
&= \operatorname{argmin}_{\alpha_0,\alpha_1} \left\{|W_i| I(W_i > 0) I(\alpha_0 + \alpha_1^T Y_i \leqslant 0) - |W_i| I(W_i \leqslant 0) I(\alpha_0 + \alpha_1^T Y_i \leqslant 0)\right\} \\
&= \operatorname{argmin}_{\alpha_0,\alpha_1} \left[|W_i| I(W_i > 0) I(\alpha_0 + \alpha_1^T Y_i \leqslant 0) - |W_i| I(W_i \leqslant 0) \left\{1 - I(\alpha_0 + \alpha_1^T Y_i \geqslant 0)\right\}\right] \\
&= \operatorname{argmin}_{\alpha_0,\alpha_1} \left\{|W_i| I(W_i > 0) I(\alpha_0 + \alpha_1^T Y_i \leqslant 0) + |W_i| I(W_i \leqslant 0) I(\alpha_0 + \alpha_1^T Y_i \geqslant 0) - |W_i| I(W_i \leqslant 0)\right\} \\
&= \operatorname{argmin}_{\alpha_0,\alpha_1} \left\{|W_i| I(W_i > 0) I(\alpha_0 + \alpha_1^T Y_i \leqslant 0) + |W_i| I(W_i \leqslant 0) I(\alpha_0 + \alpha_1^T Y_i \geqslant 0)\right\} \\
&= \operatorname{argmin}_{\alpha_0,\alpha_1} \left[|W_i| I(W_i > 0) I\left\{\operatorname{sgn}(W_i) \times (\alpha_0 + \alpha_1^T Y_i \leqslant 0)\right\} + |W_i| I(W_i \leqslant 0) I\left\{\operatorname{sgn}(W_i) \times (\alpha_0 + \alpha_1^T Y_i \leqslant 0)\right\}\right] \\
&= \operatorname{argmin}_{\alpha_0,\alpha_1} |W_i| I\left\{\operatorname{sgn}(W_i) \times (\alpha_0 + \alpha_1^T Y_i \leqslant 0)\right\} \\
&= \operatorname{argmin}_{\alpha_0,\alpha_1} |W_i| I\left\{\operatorname{sgn}(W_i) \neq \operatorname{sgn}(\alpha_0 + \alpha_1 Y_i)\right\},
\end{aligned}
$$

with subject-specific weight

$$
W_i = W_{1i} = -\left\{\frac{D_i T_i}{(N_1/N)} - \frac{D_i(1-T_i)}{(N_0/N)} + c\right\}.
$$

Second, note that the term in (3.9) that augmenting (3.8) is

$$
\sum_{i=1}^{n} \left[\frac{(1-T_i) I(\alpha_0 + \alpha_1^T Y_i \leqslant 0) + T_i I(\alpha_0 + \alpha_1^T Y_i > 0) - P\{T = I(\alpha_0 + \alpha_1^T Y_i > 0) | Y_i\}}{P\left\{T = I(\alpha_0 + \alpha_1^T Y_i > 0) | Y_i\right\}} m(Y_i; \hat\beta)\right]
$$

where $P\left\{T = I(\alpha_0 + \alpha_1^T Y_i > 0)\right\} = \pi \times I(\alpha_0 + \alpha_1^T Y_i > 0) + (1-\pi) \times I(\alpha_0 + \alpha_1^T Y_i \leqslant 0)$ for $\pi = P(T = 1)$, and $m(Y_i; \hat\beta) = \widehat{Risk}_0(Y_i) I(\alpha_0 + \alpha_1^T Y_i \leqslant 0) + \widehat{Risk}_1(Y_i) I(\alpha_0 + \alpha_1^T Y_i > 0)$, which can be further represented as

$$
\begin{aligned}
&\left\{\frac{T_i - \pi}{\pi} I(\alpha_0 + \alpha_1^T Y_i > 0) + \frac{\pi - T_i}{1-\pi} I(\alpha_0 + \alpha_1^T Y_i < 0)\right\} \times \left\{\widehat{Risk}_0 - (\widehat{Risk}_0 - \widehat{Risk}_1) \times I(\alpha_0 + \alpha_1^T Y_i > 0)\right\} \\
&\propto -\left\{\frac{\pi - T_i}{1-\pi} - \frac{T_i - \pi}{\pi}\right\} \times \widehat{Risk}_0 \times I(\alpha_0 + \alpha_1^T Y_i > 0) - \left(\frac{T_i - \pi}{\pi}\right) I(\alpha_0 + \alpha_1^T Y_i > 0) \times (\widehat{Risk}_0 - \widehat{Risk}_1) \\
&= -I(\alpha_0 + \alpha_1^T Y_i > 0) \times \left[\frac{\pi - T_i}{1-\pi} \times (\widehat{Risk}_0 + \widehat{Risk}_1)\right].
\end{aligned}
$$

Therefore, (3.9) is proportional to

$$\sum_{i=1}^{n}\left[\frac{D_i(1-T_i)I(\alpha_0+\alpha_1^TY_i>0)}{N_0}-\frac{D_iT_iI(\alpha_0+\alpha_1^TY_i>0)}{N_1}\right.$$
$$\left.-\frac{\pi-T_i}{1-\pi}\times\frac{(\widehat{Risk}_0+\widehat{Risk}_1)I(\alpha_0+\alpha_1^TY_i>0)}{N}-\frac{c\times I(\alpha_0+\alpha_1^TY_i>0)}{N}\right]$$
$$=\sum_{i=1}^{n}\left[\frac{Y_i(1-T_i)}{N_0}-\frac{Y_iT_i}{N_1}-\frac{\pi-T_i}{1-\pi}\frac{(\widehat{Risk}_0+\widehat{Risk}_1)}{N}-\frac{c}{N}\right]\times I(\alpha_0+\alpha_1^TY_i>0).$$

To maximize this is equivalent to minimizing

$$\sum_{i=1}^{n}\left[\frac{D_i(1-T_i)}{N_0}-\frac{D_iT_i}{N_1}-\frac{\pi-T_i}{1-\pi}\times\frac{(\widehat{Risk}_0+\widehat{Risk}_1)}{N}-\frac{c}{N}\right]\times I(\alpha_0+\alpha_1^TY_i\leqslant 0)$$
$$=\sum_{i=1}^{N}W_{2i}I(\alpha_0+\alpha_1^TY_i\leqslant 0)=\sum_{i=1}^{N}|W_{2i}|I\{\text{sgn}(W_{2i})\neq\text{sgn}(\alpha_0+\alpha_1^TY_i)\},$$

with

$$W_{2i}=-\left[\frac{D_iT_i}{N_1}-\frac{D_i(1-T_i)}{N_0}+\frac{\pi-T_i}{1-\pi}\times\frac{\widehat{Risk}_0+\widehat{Risk}_1}{n}+\frac{c}{N}\right].$$

Appendix D. Asymptotic Variance of Model-Based Estimators of Expected Benefit and Proof of Theorem 1

We assume the following conditions hold:

i) $\sqrt{N}(\hat{\beta}-\beta)=n^{-1/2}\sum_{i=1}^{n}I(\beta)l(\beta)_i+o_p(1)$

ii) $\rho_0-\rho_1\neq c$

iii) $E(Risk_0(\beta)-Risk_1(\beta))$ is differentiable with respect to β at the true β value

iv) $EB(c;\beta)$, $E(Risk_0(\beta)-Risk_1(\beta))_+$, $E(Risk_0(\beta))$, $E(Risk_0(\beta)+Risk_1(\beta)-1)$ are differentiable with respect to β at the true β value

We can show that when $c\neq\rho_0-\rho_1$,

$$(i)\sqrt{N}\left\{\widehat{EB}(c)-EB(c)\right\}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\Psi_{1i}+\Psi_{2i}-I(\rho_0-\rho_1>c)\left(\Psi_{3i}+\Psi_{4i}\right)\right\}+o_p(1),$$

where

$$\Psi_{1i} = \frac{\partial EB(c)}{\partial \beta} I^{-1}(\beta) l(\beta)_i, \;\; \Psi_{2i} = (\Delta_i - c)_+ - E\left\{(\Delta - c)_+\right\},$$
$$\Psi_{3i} = \frac{\partial E\{\Delta(\beta)\}}{\partial \beta} I^{-1}(\beta) l(\beta)_i, \;\; \Psi_{4i} = \Delta_i - (\rho_0 - \rho_1),$$

$$(ii) \qquad \sqrt{N}\left\{\widehat{PEB}^l(c) - PEB^l(c)\right\} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{(1-c)\left(\Psi_{5i} + \Psi_{6i}\right) - I(\rho_0 - \rho_1 > c)\left(\Psi_{3i} + \Psi_{4i}\right)\right\} + o_p(1),$$

$$(iii) \;\; \sqrt{N}\left\{\widehat{PEB}^u(c) - PEB^u(c)\right\} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left\{(1-c)\left(\Psi_{7i} + \Psi_{8i}\right) - I(\rho_0 - \rho_1 > c)\left(\Psi_{3i} + \Psi_{4i}\right)\right\} + o_p(1),$$

where
$$\Psi_{5i} = \frac{\partial E(\Delta_+)}{\partial \beta} I^{-1}(\beta) l(\beta)_i, \qquad \Psi_{6i} = (\Delta_i)_+ - E(\Delta_+),$$

$$\Psi_{7i} = \frac{\partial\left[E\{Risk_0(\beta)\} - E\{Risk_0(\beta) + Risk_1(\beta) - 1\}_+\right]}{\partial \beta} I^{-1}(\beta) l(\beta)_i,$$

$$\Psi_{8i} = Risk_{0i}(\beta) - \{Risk_{0i}(\beta) + Risk_{1i}(\beta) - 1\}_+ - \left[\rho_0 - E\{Risk_0(\beta) + Risk_1(\beta) - 1\}_+\right].$$

with $I(\beta)$ and $l(\beta)$ the information matrix and efficient influence function for $P(D|Y,T)$.

When $c \neq \rho_0 - \rho_1$, asymptotic normality of $\widehat{SEB}(c)^l$ and $\widehat{SEB}(c)^u$ then follows from Theorem 1 and the Delta method

We prove the result for $EB(c)$ and the proofs for $PEB^l(c)$ and $PEB^u(c)$ in Theorem 2 follow similar arguments.

$$\begin{aligned}
&\sqrt{N}\left\{\widehat{EB}(c) - EB(c)\right\} \\
&= \sqrt{N}\left\{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\Delta}_i - c\right)_+ - E\left(\Delta - c\right)_+\right\} - \sqrt{N}\left\{\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\Delta}_i - c\right)_+ - I(\rho_0 - \rho_1 > c)\right\} \\
&= \sqrt{N}\left\{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\Delta}_i - c\right)_+ - \frac{1}{N}\sum_{i=1}^{N}\left(\Delta_i - c\right)_+\right\} + \sqrt{N}\left\{\frac{1}{N}\sum_{i=1}^{N}\left(\Delta_i - c\right)_+ - E(\Delta - c)_+\right\} \\
&- \sqrt{N}\left\{\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\Delta}_i - c\right)_+ - (\rho_0 - \rho_1 - c)_+\right\} \\
&= \sqrt{N}\left[E\left\{\Delta(\hat{\beta}) > c\right\}_+ - E(\Delta - c)_+\right] + \sqrt{N}\left\{\frac{1}{N}\sum_{i=1}^{N}\left(\Delta_i - c\right)_+ - E(\Delta - c)_+\right\} \\
&- A + o_p(1)
\end{aligned}$$

where
$$A = \sqrt{N}\left\{\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\Delta}_i - c\right)_+ - (\rho_0-\rho_1-c)_+\right\}$$
$$= \sqrt{N}\left\{\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\Delta}_i - c\right)_+ - \left(\frac{1}{N}\sum_{i=1}^{N}\Delta_i - c\right)\times I(\rho_0-\rho_1-c>0)\right\}$$
$$+ \sqrt{N}\left\{\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\Delta}_i - c\right)\times (I\rho_0-\rho_1-c>0) - (\rho_0-\rho_1-c)_+\right\}$$

which when $\rho_0-\rho_1 \neq c$ by equi-continuity equals to

$$\sqrt{N}\times(\rho_0-\rho_1-c)\times\left\{I\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\Delta}_i - c > 0\right) - I(\rho_0-\rho_1-c>0)\right\}$$
$$+\sqrt{N}\left\{\frac{1}{N}\sum_{i=1}^{N}\hat{\Delta}_i - (\rho_0-\rho_1)\right\}\times I(\rho_0-\rho_1-c>0) + o_p(1)$$
$$=\sqrt{N}\left\{\frac{1}{N}\sum_{i=1}^{N}\hat{\Delta}_i - (\rho_0-\rho_1)\right\}\times I(\rho_0-\rho_1-c>0) + o_p(1),$$

which equals to $\sqrt{N}\left\{\sum_{i=1}^{N}\hat{\Delta}_i/N - (\rho_0 - rho_1)\right\}$ for $\rho_0-\rho_1>c$ and equals to 0 for $\rho_0-\rho_1<c$.

Appendix E. Distribution of the EB Estimator when Risk Difference Equals to Cost Ratio

When $\rho_0-\rho_1=c$, we have

$$\sqrt{N}\left\{\left(\sum_{i=1}^{N}\hat{\Delta}_i - c\right)_+ - (\rho_0-\rho_1-c)_+\right\}$$
$$=\sqrt{N}\left(\sum_{i=1}^{N}\hat{\Delta}_i - c\right)\times I\left(\sum_{i=1}^{N}\hat{\Delta}_i > c\right)$$

which equals to zero when $\sum_{i=1}^{N}\hat{\Delta}_i > c$ and $\sqrt{N}\left(\sum_{i=1}^{N}\hat{\Delta}_i - c\right)$ otherwise. Note that as $N\to\infty$, $\sqrt{N}\left(\sum_{i=1}^{N}\hat{\Delta}_i - c\right) \to^d \mathbb{Z} \equiv N\left(0, \sqrt{\rho_0(1-\rho_0)+\rho_1(1-\rho_1)}\right)$, $P\left(\sum_{i=1}^{N}\hat{\Delta}_i > c\right) \to 1/2$; and $\sqrt{N}\left\{\left(\sum_{i=1}^{N}\hat{\Delta}_i - c\right)_+ - (\rho_0-\rho_1-c)_+\right\}$ converges to $\mathbb{Z}\times I(\mathbb{Z}>0)$, which is a mixture of 0 and a $N\left(0, \sqrt{\rho_0(1-\rho_0)+\rho_1(1-\rho_1)}\right)$ distribution truncated at 0, with the probability of zero being 1/2.

Appendix F. Proof of Theorem 2 We prove the result for $EB(c)$ as the proofs for $PEB^l(c)$ and $PEB^u(c)$ are similar.

Suppose τ_N is a positive sequence of random variables converging to zero almost surely with n and satisfying $\sqrt{N}\tau_N \to \infty$ almost surely as $n \to \infty$. Define the event $\mathcal{E} \triangleq \{|\hat{\rho}_0 - \hat{\rho}_1 - c| \leqslant \tau_N\}$ then $1_{\mathcal{E}} \to 1_{\rho_0 - \rho_1 = c}$ in probability. Thus, the validity of the confidence interval follows if: (i) the projection interval provides the correct coverage when $\rho_0 - \rho_1 = c$; and (ii) the standard bootstrap confidence interval provides the correct coverage when $\rho_0 - \rho_1 \neq c$. The following theorem states the validity of the projection interval.

Theorem 3 [Projection bootstrap intervals] Assume $\Delta(Y)$ has a continuous and bounded density function. Let $\alpha, \eta \in (0, 1)$, and let c be fixed. Then,

1. $P^b\left(EB(c) \in \bigcup_{r\in\Gamma_\alpha(c)} \zeta_{EB(c),\eta}(r)\right) \geqslant 1 - \alpha - \eta + o_p(1)$;
2. $P^b\left(PEB^l(c) \in \bigcup_{r\in\Gamma_\alpha(c)} \zeta_{PEB^l(c),\eta}(r)\right) \geqslant 1 - \alpha - \eta + o_p(1)$;
3. $P^b\left(PEB^u(c) \in \bigcup_{r\in\Gamma_\alpha(c)} \zeta_{PEB^u(c),\eta}(r)\right) \geqslant 1 - \alpha - \eta + o_p(1)$.

If $E\Delta(Y) \neq c$ then the right hand side of the foregoing inequalities can be replaced with equality to $1 - \eta + o_P(1)$.

We next sketch the argument that the projection interval is valid in both (i) and (ii).

Define $EB_r(c) \triangleq E\left[\Delta(Y) - c\right]_+ - E(\Delta(Y) - c)1_{r\geqslant 0}$. We show that $\sqrt{N}(\widehat{EB}_r(c) - EB_r(c))$ and $\sqrt{N}(\widehat{EB}^b_r(c) - \widehat{EB}_r(c))$ converge to the same limiting distribution in probability. Thus, the validity of the proposed confidence intervals follows from standard arguments for the validity of projection intervals (see, for example, **?**). To simplify our proofs we assume that Y is bounded with probability one. Let $l(\beta^*)$ denote the influence function of $\sqrt{N}(\widehat{\beta} - \beta^*)$. Without loss of generality we assume $c = 0$.

For $\theta \in \mathbb{R}^{2\dim(Y)+2}$ define

$$\Delta(Y;\theta) \triangleq g^{-1}\left(\theta_0 + \theta_2^T Y\right) - g^{-1}\left(\theta_0 + \theta_1 + (\theta_2 + \theta_3)^T Y\right),$$

where g is the logit function. Note that $\Delta(Y) = \Delta(Y;\beta)$ and $\widehat{\Delta}(Y) = \Delta(Y;\widehat{\beta})$. Define $\dot{\Delta}(Y;\theta) \triangleq (d/d\theta)\Delta(Y;\theta)$, then for any compact set $\mathcal{K} \subseteq \mathbb{R}^{2\dim(Y)+2}$ the class of functions $\{||\dot{\Delta}(y;\theta)||$:

$\mathbb{R}^{\dim(Y)} \to \mathbb{R}\,,\, \theta \in \mathcal{K}\}$ is Donsker (see, for example, **?**). Write $\widehat{E}_N$ to denote expectation with respect to the empirical distribution. Then

$$\sqrt{N}(\widehat{EB}_r(0) - EB_r(0)) = \sqrt{N}\left(\widehat{E}_N\left[\widehat{\Delta}(Y)\right]_+ - \widehat{E}_N\widehat{\Delta}(Y)1_{r\geqslant 0}\right) - \sqrt{N}\left(E\left[\Delta(Y)\right]_+ - E\Delta(Y)1_{r\geqslant 0}\right),$$

which we can expand to equal

$$\sqrt{N}\widehat{E}_N\left(\left[\widehat{\Delta}(Y)\right]_+ - \left[\Delta(Y)\right]_+\right) - \widehat{E}_N\sqrt{N}\left(\widehat{\Delta}(Y) - \Delta(Y)\right)1_{r\geqslant 0} \\ + \sqrt{N}(\widehat{E}_N - E)\left(\left[\Delta(Y)\right]_+ - \Delta(Y)1_{r\geqslant 0}\right),$$

which equals

$$\widehat{E}_N\left(\left[\mathbb{Z}_N(Y) + \sqrt{N}\Delta(Y)\right]_+ - \left[\sqrt{N}\Delta(Y)\right]_+\right) \\ + \sqrt{N}(\widehat{E}_N - E)\left(\left[\Delta(Y)\right]_+ - \Delta(Y)1_{r\geqslant 0} - 1_{r\geqslant 0}E(\dot{\Delta}(Y;\beta^*)^T)l(\beta^*)\right) + o_P(1),$$

where $\mathbb{Z}_N \triangleq \sqrt{N}(\widehat{\Delta}(Y) - \Delta(Y)) = \dot{\Delta}(Y;\tilde{\beta})^T\sqrt{N}(\widehat{\beta} - \beta^*)$ for some $\tilde{\beta}$ intermediate to $\widehat{\beta}$ and β^*. We now argue that the leading term in the above display is equal to $\sqrt{N}(\widehat{E}_N - E)\dot{\Delta}(Y;\beta^*)^T l(\beta^*)1_{\Delta(Y)\geqslant 0} + o_P(1)$. The leading term in the above display is equal to

$$\widehat{E}_N\mathbb{Z}_N(Y)1_{\Delta(Y)\geqslant 0}1_{\sqrt{N}|\Delta(Y)|\geqslant|\mathbb{Z}_N(Y)|} \\ + \widehat{E}_N\left(\left[\mathbb{Z}_N(Y) + \sqrt{N}\Delta(Y)\right]_+ - \left[\sqrt{N}\Delta(Y)\right]_+\right)1_{\sqrt{N}|\Delta(Y)|\leqslant|\mathbb{Z}_N(Y)|}. \quad (0.2)$$

Note that $P\left(|\sqrt{N}\Delta(Y)| \leqslant |\mathbb{Z}_N(Y)|\right)$ is bounded above by

$$P\left(|\Delta(Y)| \leqslant \sup_{y\in\mathrm{supp}(Y)} ||\dot{\Delta}(y;\tilde{\beta})||\,||\widehat{\beta} - \beta^*||\right) \leqslant 2C \sup_{y\in\mathrm{supp}(Y)} ||\dot{\Delta}(y;\tilde{\beta})||\,||\widehat{\beta} - \beta^*|| = o_P(1),$$

where C is an upper bound on the density of $\Delta(Y)$. Using $|[a+b]_+ - [b]_+| \leqslant [a]_+$ the second term in (0.2) is bounded above in magnitude by $\widehat{E}_N\left[\mathbb{Z}_N(Y)\right]_+ 1_{\sqrt{N}|\Delta(Y)|\leqslant|\mathbb{Z}_N|} = o_P(1)$. The first term in (0.2) is equal to $\widehat{E}_N\mathbb{Z}_N(Y)1_{\Delta(Y)\geqslant 0} + o_P(1)$, which in turn is equal to $\mathbb{E}\dot{\Delta}(Y;\beta^*)^T 1_{\Delta(Y)\geqslant 0}\sqrt{N}(\widehat{E}_N - E)l(\beta^*) + o_P(1)$.

Assembling the arguments made above, it follows that

$$\sqrt{N}(\widehat{EB}_r(0) - EB_r(0)) = \nu^T \sqrt{N}(\widehat{E}_N - E) \begin{pmatrix} [\Delta(Y)]_+ \\ \Delta(Y) 1_{r \geqslant 0} \\ l(\beta^*) \end{pmatrix} + o_P(1),$$

where $\nu = (1,\, 1_{r\geqslant 0}, E\dot{\Delta}(Y;\beta^*)^T(1_{\Delta(Y)\geqslant 0} - 1_{r\geqslant 0}))^T$.

Following the same arguments, it can be shown that $\sqrt{N}(\widehat{EB}_r^b(0) - \widehat{EB}_r(0))$ equals

$$\sqrt{N}(\widehat{EB}_r^b(0) - EB_r^b(0)) = \nu^T \sqrt{N}(\widehat{E}_N - E) \begin{pmatrix} [\Delta(Y)]_+ \\ \Delta(Y) 1_{r \geqslant 0} \\ l(\beta^*) \end{pmatrix} + o_{P^b}(1),$$

where ν is defined as above and we write $r_N = o_{P^b}(1)$ to mean $P^b(|r_N| \geqslant \epsilon) = o_P(1)$ for any $\epsilon > 0$. Note that $\sqrt{N}(\widehat{\Delta}^b(Y) - \widehat{\Delta}(Y)) = \dot{\Delta}(Y;\widetilde{\beta})^T\sqrt{N}(\widehat{E}_N^b - \widehat{E}_N)I(\beta^*) + o_{P^b}(1)$ where $\widetilde{\beta}$ is intermediate to $\widehat{\beta}^b$ and $\widehat{\beta}$, and

$$\begin{aligned} P&\left(|\widehat{\Delta}(Y)| \leqslant \sup_{y\in \mathrm{supp}(Y)} ||\dot{\Delta}(y;\widetilde{\beta})||\,||\widehat{\beta}^b - \widehat{\beta}||\right) \\ &\leqslant P\left(|\Delta(Y)| \leqslant \sup_{y\in \mathrm{supp}(Y)} ||\dot{\Delta}(y;\widetilde{\beta})||\,||\widehat{\beta}^b - \widehat{\beta}|| + \sup_{y\in \mathrm{supp}(Y)} |\widehat{\Delta}(y) - \Delta(y)|\right) = o_{P^b}(1), \end{aligned}$$

where again $\widetilde{\beta}$ is intermediate to $\widehat{\beta}^b$ and $\widehat{\beta}$.

It remains to show that the bootstrap confidence interval for $\widehat{EB}(c)$ is consistent when $\rho_0 - \rho_1 \neq 0$. In the above notation this requires showing $\sqrt{N}(\widehat{EB}^b_{\hat{\rho}_0^b - \hat{\rho}_1^b}(0) - \widehat{EB}_{\hat{\rho}_0 - \hat{\rho}_1}(0))$ and $\sqrt{N}(\widehat{EB}_{\hat{\rho}_0 - \hat{\rho}_1}(0) - EB_{\rho_0 - \rho_1}(0))$ converge to the same limiting distributions in probability. However, since $\hat{\rho}_0 - \hat{\rho}_1$ is a regular, (strongly) consistent estimator of $\rho_0 - \rho_1$ and $\rho_0 - \rho_1 \neq 0$, it follows that

$$\sqrt{N}(\widehat{EB}^b_{\hat{\rho}_0^b - \hat{\rho}_1^b}(0) - \widehat{EB}_{\hat{\rho}_0 - \hat{\rho}_1}(0)) = \sqrt{N}(\widehat{EB}^b_{\rho_0 - \rho_1}(0) - \widehat{EB}_{\rho_0 - \rho_1}(0)) + o_{P^b}(1),$$

and

$$\sqrt{N}(\widehat{EB}_{\hat{\rho}_0 - \hat{\rho}_1}(0) - EB_{\rho_0 - \rho_1}(0)) = \sqrt{N}(\widehat{EB}_{\rho_0 - \rho_1}(0) - \widehat{EB}_{\rho_0 - \rho_1}(0)) + o_P(1).$$

Thus, the projection interval proof for $r = \rho_0 - \rho_1$ applies.

Appendix G. Details supporting Remark 2

Remark 2: [Locally consistent confidence interval for $EB(c)$.]

For any $\eta \in \mathbb{R}^{\dim(\beta)}$ and $r \in \mathbb{R}$ define

$$\widehat{\theta}(\eta, r) \triangleq \widehat{E}_N \Delta(Y; \widehat{\beta}_N) 1_{\Delta(Y;\eta) \geqslant 0}(1-c) - \widehat{E}_N(\Delta(Y;\eta) - c) 1_{r-c \geqslant 0},$$

$$\theta(\eta, r) \triangleq E\Delta(Y;\beta) 1_{\Delta(Y;\eta) \geqslant 0}(1-c) - E(\Delta(Y;\eta) - c) 1_{r-c \geqslant 0}.$$

Note that $\theta(\beta, E\Delta(Y;\beta)) = EB(c)$. For every fixed η, r pair it can be shown that $\sqrt{N}(\widehat{\theta}(\eta, r) - \theta(\eta, r))$ is regular, asymptotically normal, and for any $\delta \in (0,1)$ a $(1-\delta) \times 100\%$ confidence interval for $\theta(\beta, E\Delta(Y;\beta))$ can be obtained via the bootstrap. Denote such an interval by $\xi_\delta(\eta, r)$. Thus, were β^* and $E\Delta(Y;\beta)$ known, one could bootstrap $\sqrt{N}\left(\widehat{\theta}(\beta, E\Delta(Y;\beta)) - \theta(\beta, E\Delta(Y;\beta))\right)$ (but holding β and $E\Delta(Y;\beta)$ to be fixed) to obtain a valid confidence interval for $\theta(\beta, E\Delta(Y;\beta))$. Of course, neither β nor $E\Delta(Y;\beta)$ are known; however, for any $\alpha \in (0,1)$ standard methods can be used to construct a $(1-\alpha) \times 100\%$ joint confidence region for $(\beta, E\Delta(Y;\beta))$, say Γ_α. Then, it follows that

$$\bigcup_{(\eta, r) \in \Gamma_\alpha} \xi_\delta(\eta, r),$$

is a valid $(1-\delta-\alpha) \times 100\%$ confidence interval for $\theta(\beta, E\Delta(Y;\beta)) = EB(c)$. This procedure involves the union of smooth, regular, confidence intervals and is therefore also regular (i.e., locally consistent). The above interval takes a union over a larger set and is therefore potentially more conservative than the interval described in Section 4.2. On the other hand, a smooth density for $\Delta(Y)$ is no longer required (details omitted).

[*Received August 1, 2010; revised October 1, 2010; accepted for publication November 1, 2010*]