

Supplemental material to paper 'Different strategies of fitting logistic regression for positive and unlabelled data'

Paweł Teisseyre, Jan Mielniczuk and Małgorzata Łazęcka

Table 1: AUC, known c ($t = 3$)

	oracle	joint	naive	weighted
breastc	0.990	0.984	0.985	0.979
diabetes	0.813	0.804	0.806	0.804
heart-c	0.847	0.828	0.833	0.832
credit-a	0.906	0.877	0.895	0.892
credit-g	0.730	0.722	0.723	0.721
adult	0.823	0.823	0.822	0.823
vote	0.972	0.971	0.968	0.969
wdbc	0.979	0.976	0.969	0.971
spambase	0.892	0.894	0.889	0.891
avg rank	3.7	2.4	2.1	1.8

Table 2: AUC (est. c) ($t = 3$)

	oracle	joint	naive	weighted
breastc	0.990	0.985	0.985	0.983
diabetes	0.813	0.800	0.808	0.804
heart-c	0.847	0.826	0.829	0.830
credit-a	0.906	0.892	0.894	0.895
credit-g	0.730	0.720	0.723	0.714
adult	0.823	0.825	0.823	0.817
vote	0.972	0.970	0.968	0.974
wdbc	0.979	0.978	0.969	0.975
spambase	0.892	0.893	0.887	0.881
avg rank	3.7	2.3	2.0	2.0

Table 3: $|c - \hat{c}|$ ($t = 3$)

EN	TI	JM
0.085	0.052	0.034
0.225	0.185	0.103
0.160	0.164	0.056
0.125	0.129	0.137
0.290	0.268	0.204
0.277	0.252	0.157
0.048	0.107	0.115
0.083	0.073	0.033
0.225	0.302	0.046
2.3	2.2	1.4

Table 4: AUC, known c ($t = 5$)

	oracle	joint	naive	weighted
breastc	0.993	0.981	0.987	0.974
diabetes	0.821	0.805	0.808	0.805
heart-c	0.879	0.847	0.849	0.850
credit-a	0.914	0.875	0.899	0.891
credit-g	0.740	0.726	0.727	0.725
adult	0.874	0.874	0.869	0.874
vote	0.973	0.974	0.968	0.970
wdbc	0.987	0.981	0.971	0.970
spambase	0.911	0.914	0.892	0.899
rank	3.8	2.4	2.1	1.7

Table 5: AUC (est. c) ($t = 5$)

	oracle	joint	naive	weighted
breastc	0.993	0.983	0.988	0.977
diabetes	0.821	0.798	0.805	0.796
heart-c	0.879	0.843	0.850	0.853
credit-a	0.914	0.889	0.899	0.897
credit-g	0.740	0.724	0.730	0.718
adult	0.874	0.872	0.869	0.863
vote	0.973	0.972	0.968	0.977
wdbc	0.987	0.981	0.969	0.973
spambase	0.911	0.913	0.893	0.856
rank	3.8	2.2	2.2	1.8

Table 6: $|c - \hat{c}|$ ($t = 5$)

EN	TI	joint
0.060	0.064	0.030
0.234	0.169	0.071
0.138	0.121	0.043
0.125	0.130	0.317
0.287	0.261	0.143
0.244	0.214	0.059
0.044	0.088	0.024
0.099	0.068	0.033
0.189	0.267	0.033
2.4	2.3	1.2

Table 7: AUC, known c ($t = 10$)

	oracle	joint	naive	weighted
breastc	0.994	0.976	0.985	0.953
diabetes	0.827	0.801	0.804	0.799
heart-c	0.899	0.837	0.844	0.849
credit-a	0.916	0.868	0.892	0.879
credit-g	0.758	0.735	0.737	0.734
adult	0.894	0.894	0.880	0.892
vote	0.970	0.968	0.956	0.962
wdbc	0.987	0.978	0.963	0.964
spambase	0.937	0.951	0.936	0.941
avg rank	3.8	2.3	2.0	1.9

Table 8: AUC (est. c) ($t = 10$)

	oracle	joint	naive	weighted
breastc	0.994	0.979	0.983	0.953
diabetes	0.827	0.797	0.805	0.794
heart-c	0.899	0.835	0.842	0.869
credit-a	0.916	0.880	0.892	0.879
credit-g	0.758	0.726	0.734	0.724
adult	0.894	0.891	0.880	0.862
vote	0.970	0.965	0.958	0.974
wdbc	0.987	0.979	0.961	0.966
spambase	0.937	0.948	0.936	0.888
avg rank	3.8	2.3	2.2	1.7

Table 9: $|c - \hat{c}|$ ($t = 10$)

EN	TI	JM
0.063	0.031	0.018
0.229	0.153	0.079
0.140	0.111	0.106
0.192	0.095	0.316
0.279	0.233	0.237
0.237	0.223	0.061
0.070	0.063	0.049
0.074	0.042	0.039
0.144	0.271	0.021
2.8	1.9	1.3

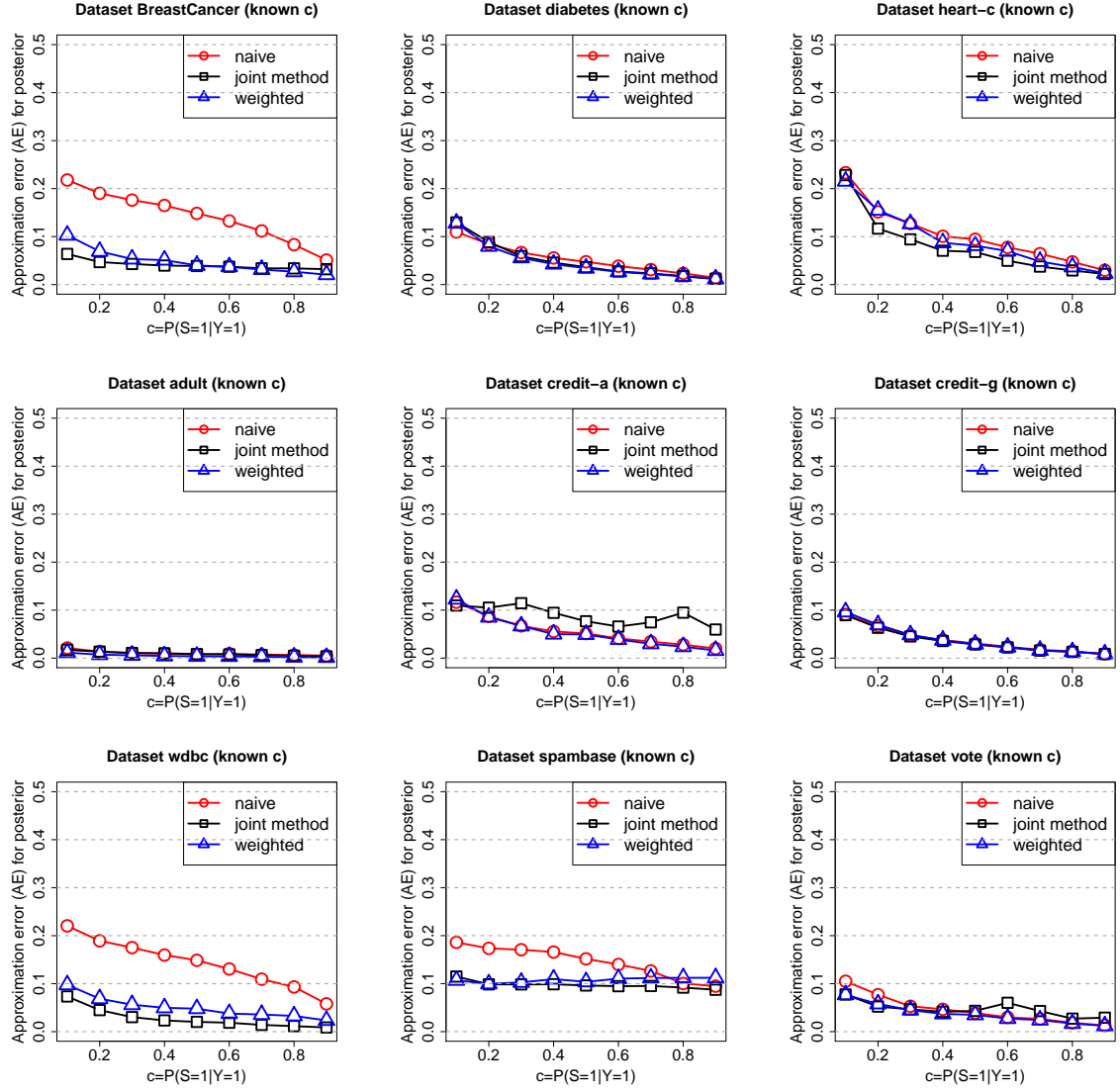


Figure 1: Approximation error for posterior wrt to c , for known c ($t = 3$).

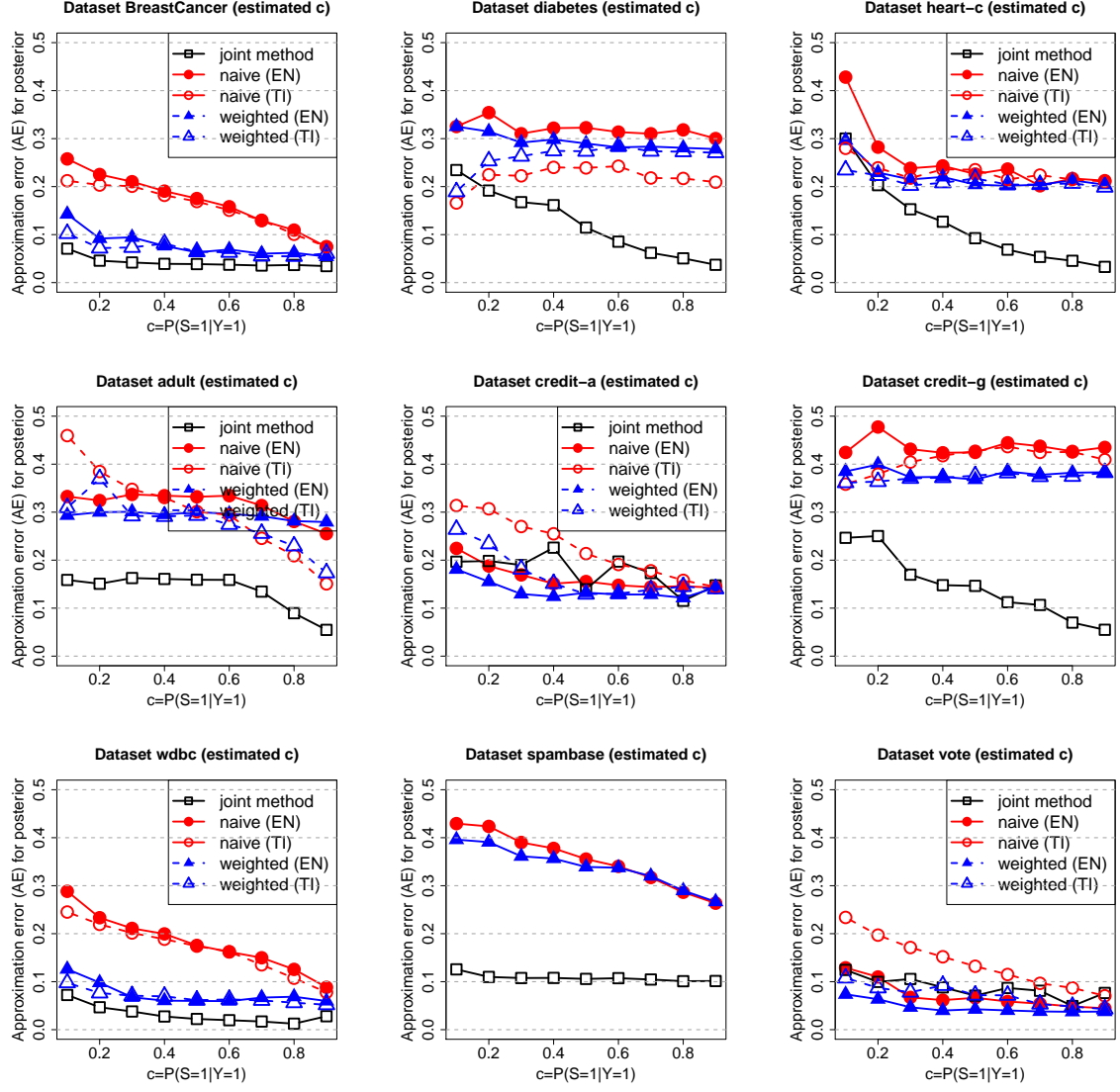


Figure 2: Approximation error for posterior wrt to c , for estimated c ($t = 3$).

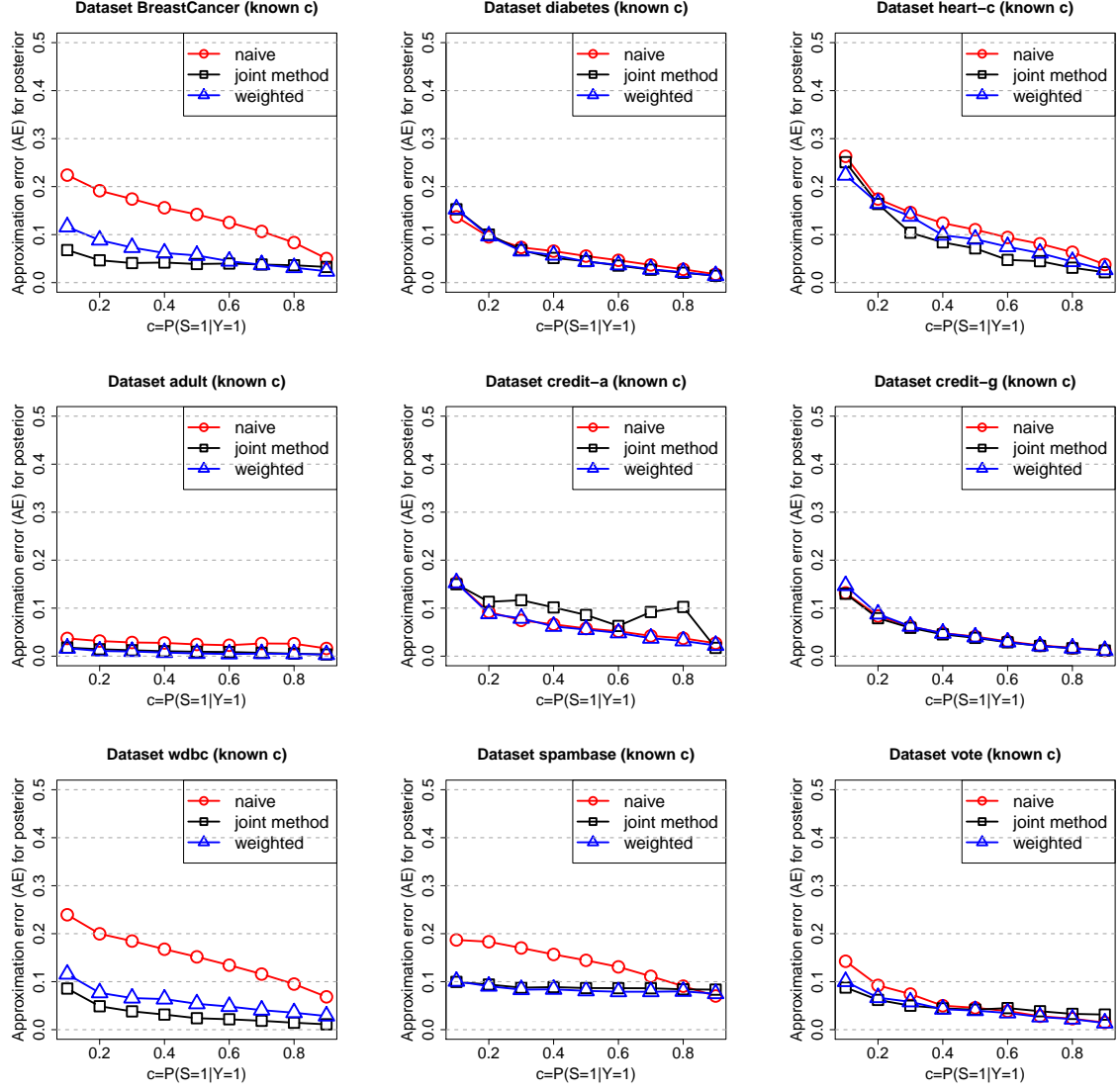


Figure 3: Approximation error for posterior wrt to c , for known c ($t = 5$).

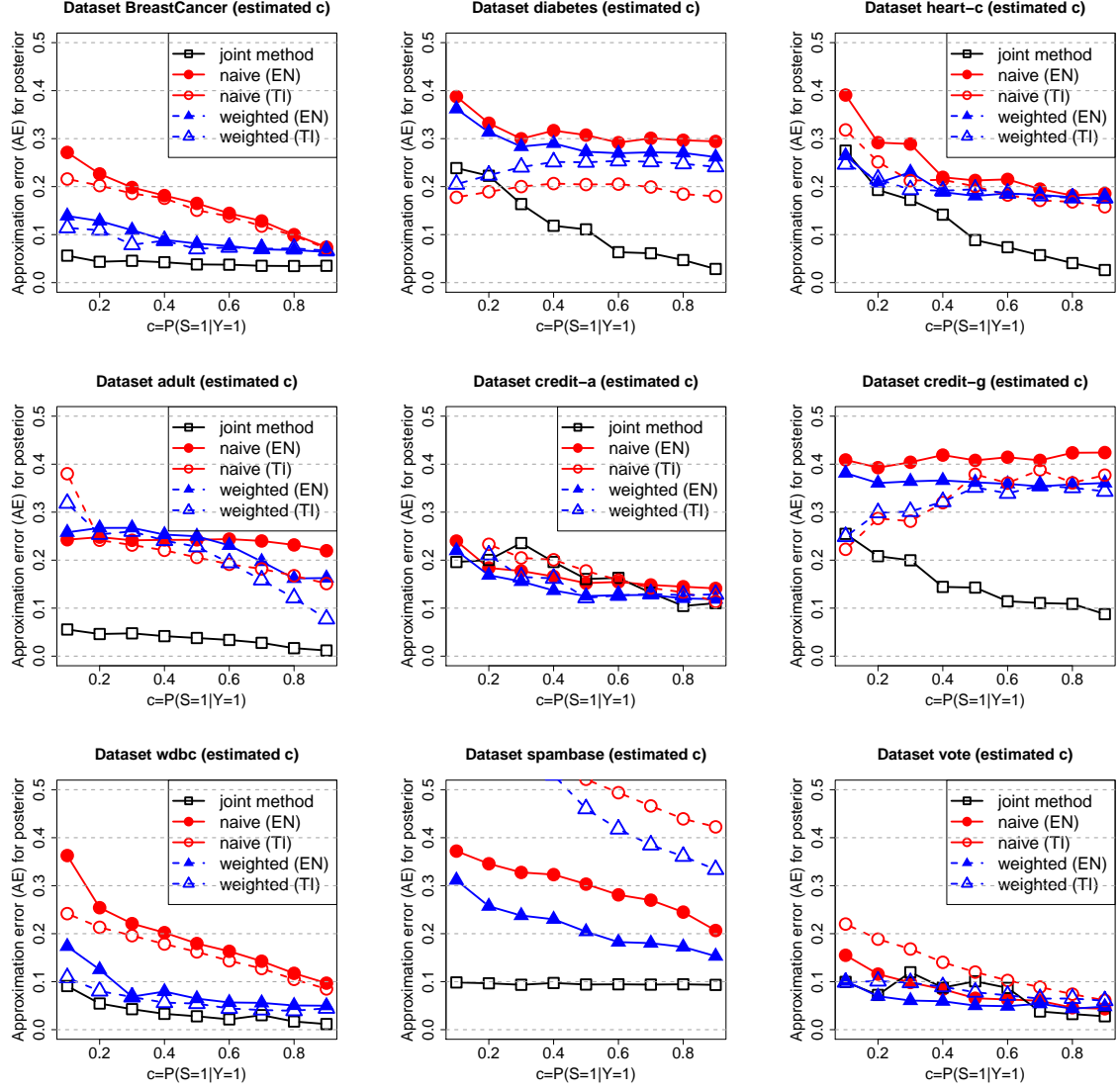


Figure 4: Approximation error for posterior wrt to c , for estimated c ($t = 5$).

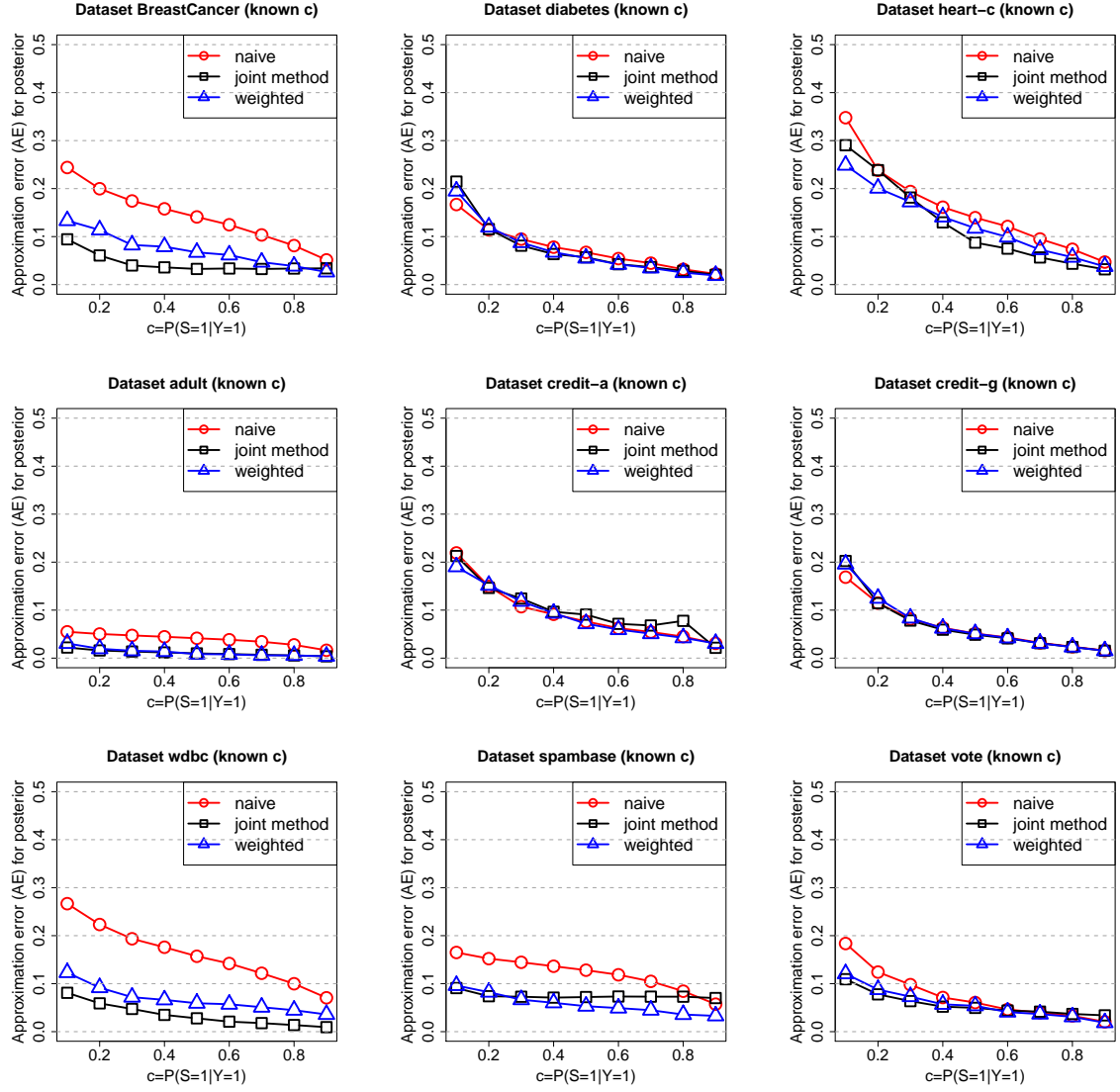


Figure 5: Approximation error for posterior wrt to c , for known c ($t = 10$).

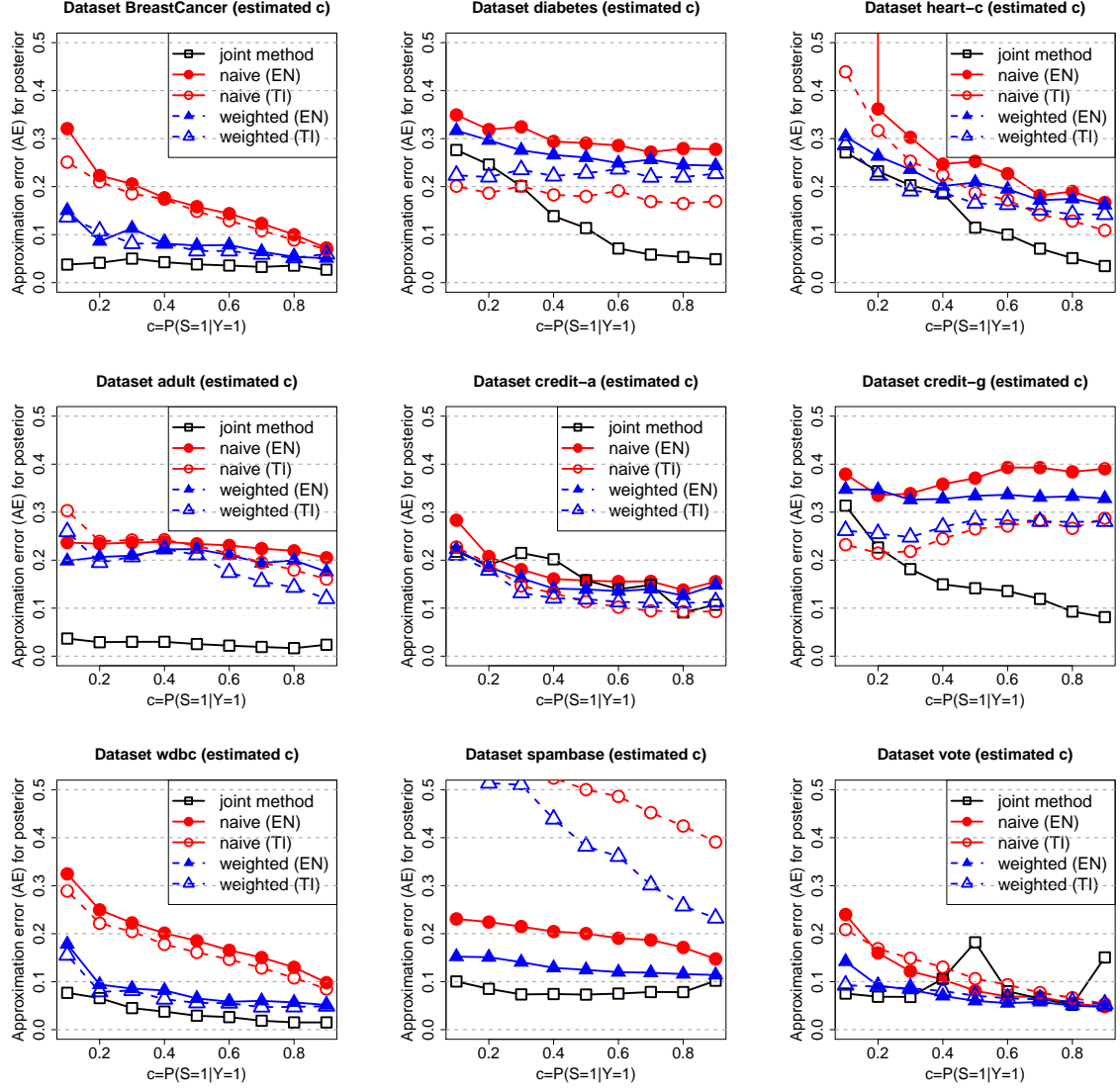


Figure 6: Approximation error for posterior wrt to c , for estimated c ($t = 10$).