Master Thesis

Simulation Results

Sanja Priselac, 01428180

October 2021

Configuration	I		II		III		IV	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
20 % positives								
LR	2.808	7.684	2.779	7.508	1.913	3.562	2.548	6.313
LR. c.s.	1.425	1.995	2.104	4.306	1.707	2.834	2.175	4.600
BY	2.807	7.677	2.773	7.480	1.858	3.361	2.548	6.314
BY. c.s.	1.425	1.998	2.171	4.591	1.661	2.687	2.154	4.512
WBY *(MCD)*	2.807	7.679	2.675	6.976	1.860	3.369	1.969	3.777
WBY. c.s. *(MCD)*	1.425	1.999	1.292	1.649	1.654	2.664	1.317	1.693
WBY *(PCDist)*	2.913	8.277	2.757	7.418	1.813	3.204	2.049	4.098
WBY. c.s. *(PCDist)*	1.543	2.346	1.410	1.964	1.503	2.205	1.133	1.265
10 % positives								
LR	4.294	17.960	3.557	12.308	2.493	6.045	3.010	8.812
LR. c.s.	2.106	4.357	2.138	4.450	2.059	4.124	2.298	5.135
BY	4.295	17.975	3.530	12.124	2.454	5.859	3.017	8.852
BY. c.s.	2.106	4.360	2.140	4.494	2.051	4.093	2.324	5.257
WBY *(MCD)*	4.296	17.988	4.176	16.995	2.468	5.926	2.752	7.372
WBY. c.s. *(MCD)*	2.107	4.364	1.989	3.895	2.058	4.122	1.758	3.011
WBY *(PCDist)*	4.492	19.670	4.308	18.100	2.442	5.800	2.840	7.854
WBY. c.s. *(PCDist)*	2.329	5.350	2.208	4.819	1.999	3.888	1.606	2.522
5 % positives								
LR	5.542	29.934	4.373	18.609	2.883	8.083	3.368	11.028
LR. c.s.	2.624	6.803	2.214	4.786	2.314	5.209	2.444	5.810
BY	5.542	29.945	4.316	18.131	2.833	7.804	3.373	11.060
BY. c.s.	2.625	6.811	2.157	4.702	2.320	5.236	2.470	5.935
WBY *(MCD)*	5.548	30.018	5.435	28.811	2.875	8.038	3.169	9.773
WBY. c.s. *(MCD)*	2.628	6.828	2.516	6.271	2.344	5.344	2.101	4.298
WBY *(PCDist)*	5.891	33.856	5.631	30.939	2.820	7.733	3.196	9.940
WBY. c.s. *(PCDist)*	2.971	8.777	2.853	8.109	2.304	5.165	2.047	4.084
1 % positives								
LR	8.499	70.912	6.679	43.583	3.421	11.379	3.920	14.942
LR. c.s.	3.891	16.288	2.564	6.709	2.692	7.050	2.778	7.507
BY	8.501	71.117	6.556	42.023	3.409	11.303	3.910	14.868
BY. c.s.	3.992	17.512	2.611	8.347	2.693	7.056	2.762	7.420
WBY *(MCD)*	8.654	74.183	8.469	70.792	3.471	11.715	3.821	14.208
WBY. c.s. *(MCD)*	4.052	18.405	3.932	17.343	2.737	7.289	2.648	6.823
WBY *(PCDist)*	11.201	126.469	9.600	91.763	3.442	11.526	3.784	13.943
WBY. c.s. *(PCDist)*	5.881	47.092	5.685	43.795	2.720	7.197	2.624	6.705

Table 1: The values of Bias and MSE obtained from 500 simulation runs for $\boldsymbol{p=2}$. The columns represent different data configurations and the rows represent the three types of estimators, both cost-sensitive (c.s.) and non-cost-sensitive, using a different imbalance proportion. Leverage detection for the initial solution in the logistic regression and the BY estimator was performed using the MCD estimator, and the WBY estimator was calculated using both methods, indicated in parentheses.

Configuration	I		II		III		IV	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
20 % positives								
LR	3.057	9.204	2.801	7.721	2.151	4.547	2.332	5.347
LR c.s.	1.676	2.804	1.645	2.688	1.994	3.909	1.919	3.625
BY	3.063	9.244	2.785	7.638	2.069	4.208	2.330	5.338
BY c.s.	1.680	2.821	1.633	2.652	1.937	3.690	1.858	3.401
WBY*(MCD)*	3.064	9.251	2.933	8.480	2.070	4.215	2.138	4.503
WBY c.s. *(MCD)*	1.682	2.827	1.549	2.411	1.935	3.682	1.554	2.386
WBY*(PCDist)*	3.094	9.438	2.963	8.656	2.052	4.143	2.148	4.548
WBY c.s. *(PCDist)*	1.712	2.933	1.582	2.514	1.904	3.567	1.526	2.307
10 % positives								
LR	4.681	21.553	4.165	17.058	2.736	7.351	2.975	8.693
LR c.s.	2.501	6.238	2.167	4.676	2.368	5.510	2.238	4.931
BY	4.691	21.665	4.135	16.820	2.678	7.045	3.023	8.981
BY c.s.	2.511	6.295	2.153	4.624	2.358	5.466	2.205	4.787
WBY*(MCD)*	4.692	21.679	4.575	20.611	2.684	7.076	2.948	8.541
WBY c.s. *(MCD)*	2.513	6.311	2.397	5.757	2.360	5.476	2.047	4.131
WBY*(PCDist)*	4.743	22.159	4.613	20.966	2.671	7.008	2.963	8.634
WBY c.s. *(PCDist)*	2.564	6.578	2.443	5.982	2.341	5.386	2.020	4.028
5 % positives								
LR	6.016	35.613	5.357	28.235	3.124	9.584	3.392	11.298
LR c.s.	3.146	9.939	2.659	7.093	2.628	6.784	2.497	6.136
BY	6.047	36.015	5.317	27.829	3.059	9.189	3.428	11.545
BY c.s.	3.173	10.136	2.661	7.119	2.634	6.817	2.489	6.095
WBY*(MCD)*	6.050	36.064	5.938	34.752	3.071	9.263	3.366	11.136
WBY c.s. *(MCD)*	3.179	10.182	3.074	9.547	2.640	6.851	2.397	5.660
WBY*(PCDist)*	6.132	37.060	5.995	35.424	3.053	9.156	3.378	11.215
WBY c.s. *(PCDist)*	3.258	10.716	3.141	9.978	2.625	6.771	2.376	5.564
1 % positives								
LR	8.891	78.301	8.028	63.740	3.635	12.971	3.984	15.591
LR c.s.	5.153	29.301	4.189	19.089	2.983	8.741	2.923	8.409
BY	9.135	83.016	8.079	64.782	3.613	12.816	3.951	15.339
BY c.s.	6.030	46.699	4.729	28.966	2.985	8.754	2.927	8.430
WBY*(MCD)*	9.205	84.456	9.081	82.186	3.637	12.983	3.913	15.050
WBY c.s. *(MCD)*	6.203	50.584	6.198	52.685	2.999	8.834	2.886	8.202
WBY*(PCDist)*	9.451	89.169	9.226	84.924	3.617	12.846	3.899	14.945
WBY c.s. *(PCDist)*	6.551	56.143	6.639	65.142	2.987	8.766	2.872	8.125

Table 2: The values of Bias and MSE obtained from 500 simulation runs for $\boldsymbol{p=10}$. The columns represent different data configurations and the rows represent the three types of estimators, both cost-sensitive (c.s.) and non-cost-sensitive, using a different imbalance proportion. Leverage detection for the initial solution in the logistic regression and the BY estimator was performed using the MCD estimator, and the WBY estimator was calculated using both methods, indicated in parentheses.

	I	II	III	IV
20 % positives				
LR	0.843 (0.009)	0.614 (0.015)	0.843 (0.009)	0.648 (0.015)
LR c.s.	0.843 (0.009)	0.620 (0.015)	0.843 (0.009)	0.659 (0.015)
BY	0.843 (0.009)	0.611 (0.015)	0.843 (0.009)	0.643 (0.015)
BY c.s.	0.843 (0.009)	0.583 (0.024)	0.843 (0.009)	0.666 (0.019)
WBY *(MCD)*	0.843 (0.009)	0.843 (0.009)	0.843 (0.009)	0.843 (0.009)
WBY c.s. *(MCD)*	0.843 (0.009)	0.843 (0.009)	0.843 (0.009)	0.843 (0.009)
WBY *(PCDist)*	0.843 (0.009)	0.843 (0.009)	0.843 (0.009)	0.843 (0.009)
WBY c.s. *(PCDist)*	0.843 (0.009)	0.843 (0.009)	0.843 (0.009)	0.843 (0.009)
10 % positives				
LR	0.865 (0.011)	0.658 (0.018)	0.865 (0.011)	0.677 (0.018)
LR c.s.	0.865 (0.011)	0.670 (0.018)	0.864 (0.011)	0.684 (0.019)
BY	0.865 (0.011)	0.654 (0.018)	0.865 (0.011)	0.675 (0.018)
BY c.s.	0.865 (0.011)	0.654 (0.030)	0.864 (0.011)	0.663 (0.037)
WBY *(MCD)*	0.865 (0.011)	0.865 (0.011)	0.865 (0.011)	0.865 (0.011)
WBY c.s. *(MCD)*	0.865 (0.011)	0.865 (0.011)	0.864 (0.011)	0.865 (0.011)
WBY *(PCDist)*	0.865 (0.011)	0.865 (0.011)	0.865 (0.011)	0.865 (0.011)
WBY c.s. *(PCDist)*	0.865 (0.011)	0.865 (0.011)	0.864 (0.011)	0.865 (0.011)
5 % positives				
LR	0.884 (0.013)	0.697 (0.023)	0.883 (0.013)	0.697 (0.024)
LR c.s.	0.884 (0.013)	0.711 (0.023)	0.883 (0.013)	0.687 (0.026)
BY	0.884 (0.013)	0.693 (0.023)	0.884 (0.013)	0.700 (0.023)
BY c.s.	0.884 (0.013)	0.708 (0.042)	0.883 (0.013)	0.662 (0.046)
WBY *(MCD)*	0.884 (0.013)	0.884 (0.013)	0.883 (0.013)	0.884 (0.013)
WBY c.s. *(MCD)*	0.884 (0.013)	0.884 (0.013)	0.882 (0.014)	0.883 (0.013)
WBY *(PCDist)*	0.884 (0.013)	0.884 (0.013)	0.884 (0.013)	0.884 (0.013)
WBY c.s. *(PCDist)*	0.884 (0.013)	0.884 (0.013)	0.882 (0.014)	0.883 (0.013)
1 % positives				
LR	0.923 (0.026)	0.780 (0.048)	0.894 (0.076)	0.442 (0.179)
LR c.s.	0.922 (0.026)	0.791 (0.047)	0.889 (0.084)	0.321 (0.180)
BY	0.923 (0.026)	0.777 (0.048)	0.898 (0.072)	0.524 (0.161)
BY c.s.	0.921 (0.027)	0.823 (0.074)	0.890 (0.083)	0.396 (0.188)
WBY *(MCD)*	0.922 (0.026)	0.923 (0.026)	0.824 (0.172)	0.888 (0.107)
WBY c.s. *(MCD)*	0.921 (0.027)	0.921 (0.027)	0.813 (0.180)	0.876 (0.117)
WBY *(PCDist)*	0.923 (0.026)	0.923 (0.026)	0.836 (0.168)	0.890 (0.117)
WBY c.s. *(PCDist)*	0.921 (0.027)	0.920 (0.027)	0.823 (0.177)	0.878 (0.120)

Table 3: The mean values of the Gini index with the standard deviations in parenthesis obtained from 500 simulation runs for $\boldsymbol{p=2}$. The columns represent different data configurations and the rows represent the three types of estimators, both cost-sensitive (c.s.) and non-cost-sensitive, using a different imbalance proportion. Leverage detection for the initial solution in the logistic regression and the BY estimator was performed using the MCD estimator, and the WBY estimator was calculated using both methods, indicated in parentheses.

	I	II	III	IV
20 % positives				
LR	0.867 (0.008)	0.814 (0.010)	0.865 (0.008)	0.812 (0.010)
LR c.s.	0.867 (0.008)	0.812 (0.010)	0.864 (0.008)	0.806 (0.010)
BY	0.867 (0.008)	0.814 (0.010)	0.866 (0.008)	0.815 (0.010)
BY c.s.	0.867 (0.008)	0.812 (0.010)	0.865 (0.008)	0.810 (0.010)
WBY *(MCD)*	0.867 (0.008)	0.867 (0.008)	0.866 (0.008)	0.867 (0.008)
WBY c.s. *(MCD)*	0.867 (0.008)	0.867 (0.008)	0.865 (0.008)	0.866 (0.008)
WBY *(PCDist)*	0.867 (0.008)	0.867 (0.008)	0.866 (0.008)	0.866 (0.008)
WBY c.s. *(PCDist)*	0.867 (0.008)	0.867 (0.008)	0.865 (0.008)	0.866 (0.008)
10 % positives				
LR	0.887 (0.009)	0.843 (0.012)	0.884 (0.010)	0.837 (0.012)
LR c.s.	0.887 (0.009)	0.841 (0.012)	0.881 (0.011)	0.825 (0.013)
BY	0.887 (0.009)	0.844 (0.012)	0.885 (0.010)	0.843 (0.012)
BY c.s.	0.887 (0.009)	0.841 (0.012)	0.881 (0.010)	0.830 (0.013)
WBY *(MCD)*	0.887 (0.009)	0.887 (0.009)	0.885 (0.010)	0.886 (0.009)
WBY c.s. *(MCD)*	0.887 (0.009)	0.886 (0.009)	0.881 (0.011)	0.883 (0.010)
WBY *(PCDist)*	0.887 (0.009)	0.887 (0.009)	0.885 (0.010)	0.886 (0.009)
WBY c.s. *(PCDist)*	0.886 (0.009)	0.886 (0.009)	0.881 (0.010)	0.883 (0.010)
5 % positives				
LR	0.906 (0.012)	0.870 (0.014)	0.898 (0.013)	0.856 (0.015)
LR c.s.	0.905 (0.012)	0.868 (0.014)	0.892 (0.014)	0.836 (0.018)
BY	0.906 (0.012)	0.870 (0.014)	0.902 (0.012)	0.866 (0.014)
BY c.s.	0.905 (0.012)	0.867 (0.014)	0.892 (0.014)	0.842 (0.017)
WBY *(MCD)*	0.906 (0.012)	0.906 (0.012)	0.901 (0.012)	0.904 (0.012)
WBY c.s. *(MCD)*	0.904 (0.012)	0.904 (0.012)	0.890 (0.015)	0.895 (0.014)
WBY *(PCDist)*	0.906 (0.012)	0.906 (0.012)	0.901 (0.012)	0.904 (0.012)
WBY c.s. *(PCDist)*	0.904 (0.012)	0.904 (0.012)	0.891 (0.015)	0.895 (0.014)
1 % positives				
LR	0.935 (0.019)	0.911 (0.023)	0.848 (0.074)	0.812 (0.061)
LR c.s.	0.927 (0.021)	0.903 (0.026)	0.824 (0.085)	0.746 (0.080)
BY	0.934 (0.019)	0.910 (0.023)	0.867 (0.064)	0.852 (0.047)
BY c.s.	0.925 (0.023)	0.901 (0.027)	0.822 (0.086)	0.756 (0.079)
WBY *(MCD)*	0.933 (0.019)	0.933 (0.019)	0.843 (0.084)	0.903 (0.036)
WBY c.s. *(MCD)*	0.924 (0.023)	0.923 (0.023)	0.795 (0.102)	0.850 (0.062)
WBY *(PCDist)*	0.933 (0.019)	0.933 (0.019)	0.856 (0.077)	0.907 (0.033)
WBY c.s. *(PCDist)*	0.923 (0.023)	0.923 (0.023)	0.808 (0.098)	0.855 (0.059)

Table 4: The mean values of the Gini index with the standard deviations in parenthesis obtained from 500 simulation runs for ***p=10***. The columns represent different data configurations and the rows represent the three types of estimators, both cost-sensitive (c.s.) and non-cost-sensitive, using a different imbalance proportion. Leverage detection for the initial solution in the logistic regression and the BY estimator was performed using the MCD estimator, and the WBY estimator was calculated using both methods, indicated in parentheses.