Spline-Rule Ensemble Classifiers with Structured Sparsity Regularization for Interpretable Customer Churn Modeling: Online Appendix

Koen W. De Bock[1] and Arno De Caigny[2,3]

[1] Audencia Business School, 8 Route de la Jonelière, F-44312, Nantes, France

[2] IESEG School of Management, 3 Rue de la Digue, F-59000, Lille, France

[3] LEM-CNRS 9221, 3 Rue de la Digue, F-59000, Lille, France

E-mail addresses: kdebock@audencia.com (Koen W. De Bock), a.de-caigny@ieseg.fr (Arno De Caigny)

Declaration of Interest: none

Corresponding author: Koen W. De Bock, Audencia Business School, 8 Route de la Jonelière, F-44312, Nantes, France. E-mail address: kdebock@audencia.com, Tel.: +33 2 40 37 34 34

Empirical Evaluation of Spline Type Choice on SRE and SRE-SGL Performance

The choice for penalized cubic regression splines in SRE-SGL is inspired on the study in which SRE was first presented (De Bock, 2017). In this online appendix, we further investigate the adoption of alternative spline types as well as shrinkage-enforcing spline variants. A priori there are arguments both in favor of, and against this modification in search of increased model performance. On the one hand: SRE (and SRE-SGL) apply shrinkage in their final step: the application of a regularized regression to the pool of candidate terms. This regularization ensures model/term selection optimizing loss over the *full* set of candidate terms. Preliminary shrinkage, pursued at the time of spline estimation, effectively reduces the set of candidate terms, limiting the playing field of subsequent regularization. On the other hand, it is known that simultaneous shrinkage and curvature penalization could lead to better performance in GAMs.

Specifically, we conduct an experimental comparison of five variations of SRE and SRE-SGL implemented with the following spline types:

- Penalized cubic regression splines (identified using the suffix *cr*)
- Penalized cubic regression splines with shrinkage (identified using the suffix *cs*)
- Penalized thin plate regression splines (identified using the suffix *tp*)
- P-splines (identified using the suffix *ps*)
- Penalized thin plate regression splines with shrinkage (identified using the suffix *ts*)

Experimental conditions and data sets are fully described in the main manuscript. Results of this additional benchmarking experiment are presented in the Tables (A-D) below. Average ranks and adjusted p-values are reported in Tables (E-F). Using the same statistical analysis as fully disclosed in the benchmarking experiment in the paper, we could not detect any significant differences between the penalized cubic regression splines (*cr*) on the one hand and the variations adopting alternative spline types on the other.

Data set	*SRE - cr*	*SRE - cs*	*SRE - ps*	*SRE - ts*	*SRE - tp*
Ds1	0.885 (0.003)	0.879 (0.002)	0.884 (0.002)	0.881 (0.003)	0.882 (0.003)
Ds2	0.862 (0.001)	0.861 (0.001)	0.859 (0.001)	0.854 (0.001)	0.855 (0.001)
Ds3	0.653 (0.008)	0.652 (0.007)	0.648 (0.005)	0.660 (0.006)	0.660 (0.006)
Ds4	0.843 (0.002)	0.842 (0.001)	0.844 (0.002)	0.839 (0.001)	0.840 (0.002)
Ds5	0.791 (0.002)	0.792 (0.002)	0.788 (0.001)	0.795 (0.001)	0.792 (0.001)
Ds6	0.789 (0.001)	0.781 (0.002)	0.789 (0.002)	0.785 (0.002)	0.784 (0.002)
Ds7	0.716 (0.005)	0.715 (0.004)	0.720 (0.005)	0.714 (0.005)	0.714 (0.005)
Ds8	0.831 (0.003)	0.829 (0.003)	0.821 (0.002)	0.831 (0.002)	0.830 (0.002)
Ds9	0.626 (0.005)	0.628 (0.004)	0.621 (0.005)	0.611 (0.004)	0.612 (0.004)
Ds10	0.620 (0.004)	0.618 (0.005)	0.621 (0.004)	0.609 (0.004)	0.608 (0.004)
Ds11	0.725 (0.014)	0.722 (0.012)	0.730 (0.013)	0.721 (0.012)	0.722 (0.011)
Ds12	0.816 (0.002)	0.817 (0.002)	0.814 (0.003)	0.820 (0.003)	0.819 (0.003)
Ds13	0.863 (0.004)	0.864 (0.003)	0.859 (0.004)	0.862 (0.003)	0.862 (0.003)
Ds14	0.756 (0.012)	0.754 (0.011)	0.753 (0.012)	0.758 (0.012)	0.758 (0.012)

Table A: Average AUC - results for SRE variants

Data set	*SRE - cr*	*SRE - cs*	*SRE - ps*	*SRE - ts*	*SRE - tp*
Ds1	5.515 (0.249)	5.560 (0.242)	5.305 (0.210)	4.988 (0.193)	5.002 (0.205)
Ds2	5.177 (0.047)	5.050 (0.052)	5.258 (0.055)	5.351 (0.044)	5.005 (0.046)
Ds3	2.053 (0.153)	2.005 (0.150)	2.152 (0.168)	2.105 (0.161)	2.053 (0.158)
Ds4	4.407 (0.037)	4.102 (0.040)	4.360 (0.045)	4.485 (0.040)	4.522 (0.039)
Ds5	3.701 (0.072)	3.801 (0.065)	3.652 (0.080)	3.789 (0.082)	3.659 (0.078)
Ds6	3.593 (0.063)	3.410 (0.065)	3.337 (0.052)	3.466 (0.063)	3.687 (0.064)
Ds7	2.333 (0.285)	2.315 (0.287)	2.250 (0.197)	2.346 (0.211)	2.346 (0.208)
Ds8	4.564 (0.122)	4.488 (0.110)	4.378 (0.140)	4.322 (0.122)	4.488 (0.118)
Ds9	1.585 (0.027)	1.655 (0.030)	1.629 (0.034)	1.499 (0.041)	1.629 (0.044)
Ds10	1.468 (0.616)	1.501 (0.430)	1.444 (0.510)	1.318 (0.055)	1.408 (0.620)
Ds11	3.322 (0.185)	3.288 (0.170)	3.322 (0.162)	3.488 (0.165)	3.440 (0.172)
Ds12	2.782 (0.058)	2.882 (0.078)	2.533 (0.066)	2.436 (0.081)	2.533 (0.077)
Ds13	5.246 (0.077)	5.273 (0.095)	5.336 (0.080)	5.011 (0.073)	5.098 (0.076)
Ds14	2.072 (0.430)	2.052 (0.377)	2.158 (0.412)	2.358 (0.442)	2.255 (0.444)

Table B: Average TDL - results for SRE variants

Data set	*SRE - cr*	*SRE - cs*	*SRE - ps*	*SRE - ts*	*SRE - tp*
Ds1	0.883 (0.003)	0.877 (0.003)	0.884 (0.002)	0.882 (0.003)	0.883 (0.003)
Ds2	0.862 (0.001)	0.862 (0.002)	0.858 (0.002)	0.863 (0.001)	0.862 (0.001)
Ds3	0.656 (0.006)	0.654 (0.004)	0.648 (0.005)	0.661 (0.004)	0.658 (0.005)
Ds4	0.843 (0.002)	0.843 (0.003)	0.846 (0.003)	0.842 (0.002)	0.842 (0.002)
Ds5	0.790 (0.002)	0.792 (0.002)	0.789 (0.001)	0.794 (0.001)	0.793 (0.002)
Ds6	0.789 (0.001)	0.828 (0.002)	0.822 (0.001)	0.832 (0.001)	0.831 (0.001)
Ds7	0.711 (0.006)	0.710 (0.004)	0.715 (0.005)	0.709 (0.004)	0.711 (0.004)
Ds8	0.830 (0.004)	0.785 (0.003)	0.787 (0.004)	0.785 (0.004)	0.784 (0.003)
Ds9	0.625 (0.004)	0.624 (0.002)	0.623 (0.003)	0.619 (0.003)	0.619 (0.003)
Ds10	0.623 (0.004)	0.628 (0.005)	0.620 (0.004)	0.619 (0.005)	0.619 (0.005)
Ds11	0.728 (0.017)	0.729 (0.015)	0.729 (0.012)	0.724 (0.013)	0.724 (0.013)
Ds12	0.815 (0.002)	0.816 (0.002)	0.813 (0.001)	0.814 (0.002)	0.814 (0.002)
Ds13	0.863 (0.006)	0.864 (0.004)	0.858 (0.005)	0.862 (0.004)	0.863 (0.004)
Ds14	0.761 (0.010)	0.755 (0.009)	0.758 (0.010)	0.759 (0.009)	0.759 (0.009)

Table C: Average AUC results for SRE-SGL variants

Data set	*SRE - cr*	*SRE - cs*	*SRE - ps*	*SRE - ts*	*SRE - tp*
Ds1	5.487 (0.272)	5.268 (0.252)	5.551 (0.271)	5.502 (0.211)	5.218 (0.252)
Ds2	5.149 (0.049)	5.306 (0.051)	5.149 (0.055)	5.009 (0.049)	5.015 (0.044)
Ds3	2.097 (0.127)	2.213 (0.123)	2.255 (0.114)	2.117 (0.114)	2.005 (0.127)
Ds4	4.404 (0.038)	4.205 (0.042)	4.008 (0.036)	4.265 (0.042)	4.489 (0.041)
Ds5	3.705 (0.085)	3.591 (0.088)	3.408 (0.085)	3.520 (0.093)	3.782 (0.085)
Ds6	3.568 (0.067)	4.655 (0.064)	4.441 (0.064)	4.707 (0.062)	4.655 (0.061)
Ds7	2.347 (0.285)	2.456 (0.279)	2.200 (0.238)	2.551 (0.240)	2.504 (0.211)
Ds8	4.560 (0.114)	3.780 (0.113)	3.588 (0.129)	3.307 (0.114)	3.442 (0.097)
Ds9	1.587 (0.027)	1.506 (0.026)	1.503 (0.025)	1.400 (0.028)	1.355 (0.028)
Ds10	1.628 (0.283)	1.488 (0.220)	1.422 (0.275)	1.467 (0.244)	1.722 (0.220)
Ds11	3.537 (0.246)	3.442 (0.236)	3.242 (0.226)	3.688 (0.244)	3.255 (0.228)
Ds12	2.777 (0.059)	2.573 (0.068)	2.805 (0.060)	2.892 (0.061)	2.777 (0.059)
Ds13	5.250 (0.091)	5.407 (0.082)	5.428 (0.086)	5.351 (0.085)	5.407 (0.088)
Ds14	2.231 (0.090)	2.123 (0.091)	2.105 (0.094)	2.057 (0.90)	2.289 (0.091)

Table D: Average TDL results for SRE-SGL variants

		Metric	
Algorithm role	**Algorithm**	**AUC**	**TDL**
Control	SRE - cr	**2.214**	**2.786**
Benchmarks	SRE - cs	3.143 (0.292)	2.964 (0.999)
	SRE - ps	3.107 (0.292)	3.250 (0.999)
	SRE - ts	3.250 (0.292)	3.107 (0.999)
	SRE - tp	3.286 (0.292)	2.893 (0.999)

[1] Lower average ranks indicate better performance. The best performing algorithm is indicated in bold. The adjusted p-value for Holm post-hoc test is shown between brackets.

Table E: Average SRE variations ranks across data sets for different performance measures and significance test results

		Metric	
Algorithm role	**Algorithm**	**AUC**	**TDL**
Control	SRE-SGL - cr	**2.357**	**2.714**
Benchmarks	SRE-SGL - cs	2.893 (0.427)	2.857 (0.999)
	SRE-SGL - ps	3.321 (0.427)	3.393 (0.999)
	SRE-SGL - ts	3.231 (0.427)	3.071 (0.999)
	SRE-SGL - tp	3.250 (0.427)	2.964 (0.999)

[1] Lower average ranks indicate better performance. The best performing algorithm is indicated in bold. The adjusted p-value for Holm post-hoc test is shown between brackets.

Table F: Average SRE-SGL variations ranks across data sets for different performance measures and significance test results