Supplementary Materials for "Likelihood based inference for censored linear regression models with scale mixtures of skew-normal distributions"

Thalita do Bem Mattos[a], Aldo M. Garay[b] and Victor H. Lachos[a]

September 29, 2017

A Complementary results of the simulation study 1

In this appendix, we present the results of the simulation study 1 for different levels of censoring: $p = 0\%,\ 20\%$ and 35%.

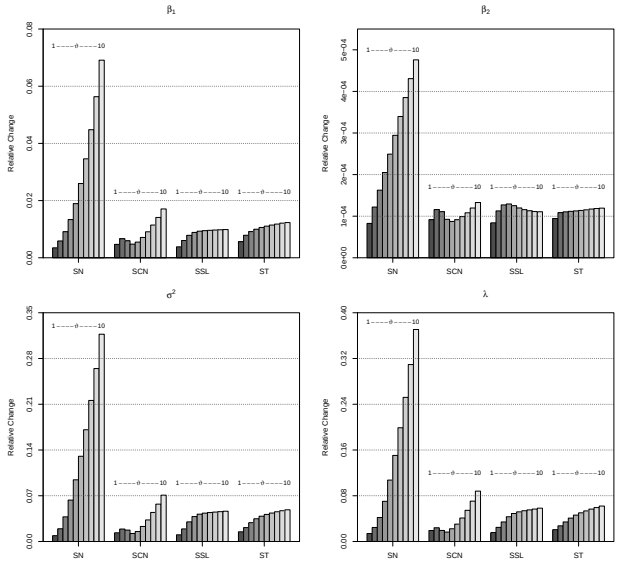

Figure A.1: Simulation study 1. Average relative changes of estimates for different perturbations ϑ and $p = 0\%$.

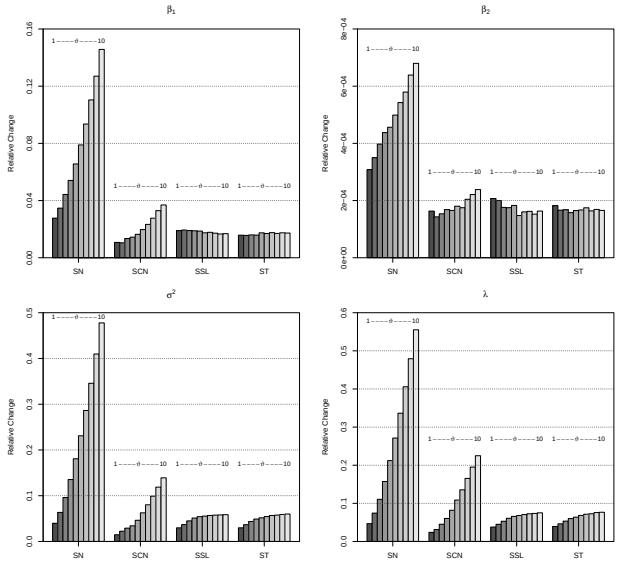

Figure A.2: Simulation study 1. Average relative changes of estimates for different perturbations ϑ and $p = 20\%$.

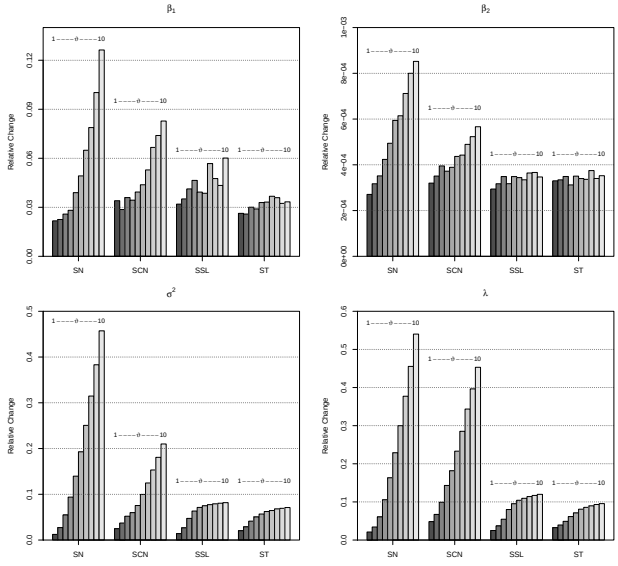

Figure A.3: Simulation study 1. Average relative changes of estimates for different perturbations ϑ and $p = 35\%$.

B Complementary results of the simulation study 2

Here we show the Bias and MSE of parameters $\boldsymbol{\theta}$, for the censoring level $p = 20\%$ and 35%, respectively.

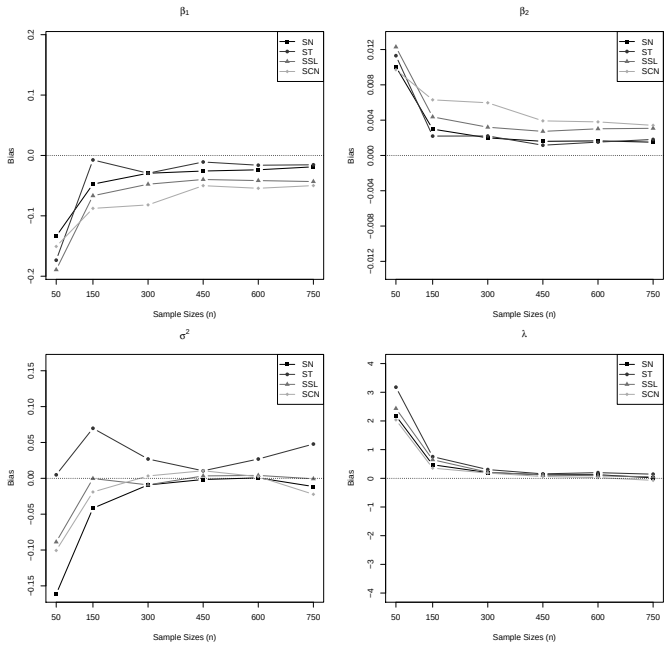

Figure B.1: Simulation study 2. Bias of parameters β_1, β_2, σ^2 and λ for SMSN-models with $p = 20\%$.

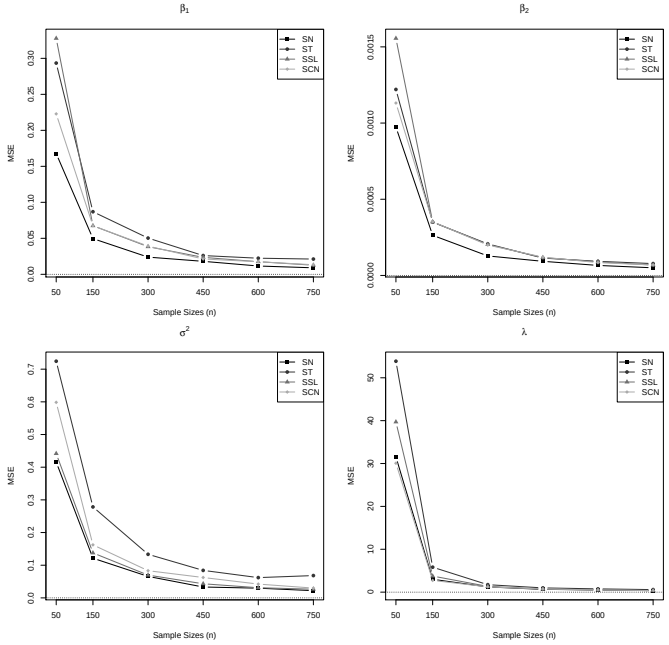

Figure B.2: Simulation study 2. MSE of parameters β_1, β_2, σ^2 and λ for SMSN-models with $p = 20\%$.

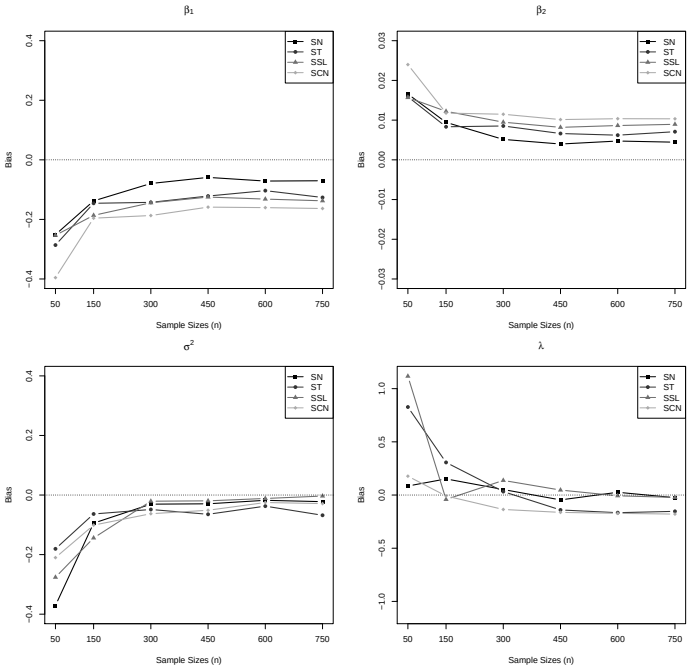

Figure B.3: Simulation study 2. Bias of parameters β_1, β_2, σ^2 and λ for SMSN-models with $p = 35\%$.

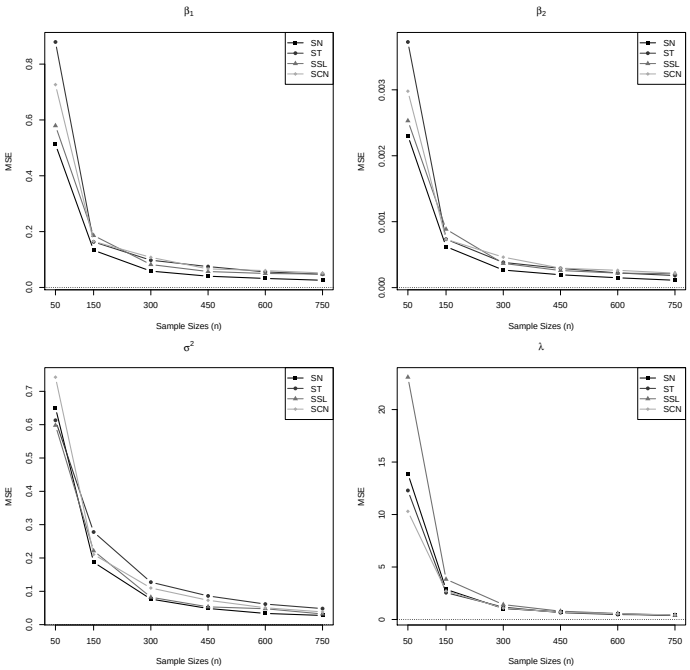

Figure B.4: Simulation study 2. MSE of parameters β_1, β_2, σ^2 and λ for SMSN-models with $p = 35\%$.

C Case studies under SMN-CR models

In this Appendix, we present in Table C.1 and C.2, the comparison among the SMN-CR models, considering the *wage rate* dataset and the *stellar abundances* dataset, respectively.

Table C.1: Wage rate data. Values of selected model selection criteria for SMN-CR models.

Criteria	N-CR	T-CR	CN-CR	SL-CR
log-likelihood	-1481.6550	-1440.1450	-1432.0850	-1436.2860
AIC	2975.3110	2894.2910	2880.1710	2886.5730
BIC	3003.0550	2926.6590	2917.1630	2918.9410
EDC	2996.2400	2918.7080	2908.0760	2910.9900

Table C.2: Stellar abundances data. Values of selected model selection criteria for SMN-CR models.

Criteria	N-CR	T-CR	CN-CR	SL-CR
log-likelihood	-38.2808	-6.9309	-17.1163	-14.6924
AIC	82.5615	21.8617	44.2326	37.3848
BIC	89.2200	30.7397	55.3302	46.2629
EDC	81.5092	20.4587	42.4789	35.9818

D Complementary results of the simulation study

In this appendix, we present a simulation study to evaluate the behavior of the proposed SAEM algorithm in the presence of random explanatory variables. We use a similar design to that considered in simulation study 2 (Section 5.2), the only difference being that the values of variable $x_i, i = 1, \ldots, n$, were generated independently from a uniform distribution in the interval (2,20) and new values were generated for each iteration. Were considered 500 replicates, the three SMSN-CR models with censoring levels $p = \{8\%,\ 20\%,\ 35\%\}$ and samples sizes fixed at $n = \{50,\ 150,\ 300,\ 450,\ 600\}$.

Thus, under this scenario, we evaluated the parameter estimates $\widehat{\boldsymbol{\theta}} = \left(\widehat{\beta}_1, \widehat{\beta}_2\right)$ obtained considering two methods: (i) our proposed SAEM algorithm, for SMSN-CR models, and (ii) the linear quantile regression model, by using the R package *lqr*$(\cdot)$, developed by [1]. It is important to note that in this second approach we considered the original dependence variable generated, without considering the presence of censored values.

We chose the *lqr*$(\cdot)$ package, as a good alternative that allows estimating the parameters of a robust linear quantile regression model, considering a new family of zero-quantile distributions for the error term, which includes skewed versions of the normal, Student-t, contaminated normal distribution, among others.

In addition to observations described above, in the text, we observe that in general when we used the *lqr*(·) package to estimate the parameters, the bias and MSE tended to zero when n increased, with the exception of $\widehat{\beta}_1$ for the skew-t model.

Table D.1: Simulation Study. Bias and MSE for the parameters β_1 and β_2, obtained by using our proposed SAEM algorithm and the *lqr*(·) package, with $p = 8\%$.

Bias							
		SN		**ST**		**SCN**	
		`lqr()`	SMSN-CR	`lqr()`	SMSN-CR	`lqr()`	SMSN-CR
$\widehat{\beta}_1$	$n = 50$	0.0166	-0.0659	-0.3716	-0.0938	-0.0414	-0.1184
	$n = 150$	0.0085	-0.0248	-0.3959	-0.0442	-0.0200	-0.0365
	$n = 300$	-0.0068	-0.0250	-0.3934	-0.0220	-0.0001	-0.0262
	$n = 450$	-0.0066	-0.0207	-0.3951	-0.0135	-0.0057	-0.0235
	$n = 600$	0.0008	-0.008	-0.3983	-0.0178	-0.0080	-0.0143
$\widehat{\beta}_2$	$n = 50$	-0.0018	0.0040	-0.0008	0.0067	0.0032	0.00757
	$n = 150$	-0.0004	0.0020	-0.0008	0.0027	0.0021	0.00271
	$n = 300$	0.0008	0.0020	-0.0004	0.0017	0.0004	0.00224
	$n = 450$	0.0005	0.0015	-0.0002	0.0009	< 0.0001	0.00127
	$n = 600$	0.0001	0.0008	-0.0001	0.0010	0.0003	0.00098
MSE							
		SN		**ST**		**SCN**	
		`lqr()`	SMSN-CR	`lqr()`	SMSN-CR	`lqr()`	SMSN-CR
$\widehat{\beta}_1$	$n = 50$	0.0882	0.0925	0.3076	0.1573	0.2237	0.1470
	$n = 150$	0.0325	0.0311	0.2024	0.0514	0.0760	0.0345
	$n = 300$	0.0140	0.0147	0.1799	0.0243	0.0443	0.0190
	$n = 450$	0.0089	0.0092	0.1725	0.0168	0.0235	0.0129
	$n = 600$	0.0072	0.0066	0.1700	0.0116	0.0173	0.0098
$\widehat{\beta}_2$	$n = 50$	0.0006	0.0006	0.0011	0.0008	0.0015	0.0008
	$n = 150$	0.0002	0.0002	0.0003	0.0002	0.0005	0.0002
	$n = 300$	0.0001	0.0001	0.0002	0.0001	0.0003	0.0001
	$n = 450$	0.0001	0.0001	0.0001	0.0001	0.0002	0.0001
	$n = 600$	0.0001	< 0.0001	0.0001	0.0001	0.0001	0.0001

Table D.2: Simulation Study. Bias and MSE for the parameters β_1 and β_2, obtained by using our proposed SAEM algorithm and the *lqr*(·) package, with $p = 20\%$.

Bias							
		SN		**ST**		**SCN**	
		`lqr()`	SMSN-CR	`lqr()`	SMSN-CR	`lqr()`	SMSN-CR
$\widehat{\beta}_1$	$n = 50$	-0.0105	-0.1169	-0.4116	-0.0656	0.0367	-0.1197
	$n = 150$	-0.0128	-0.0549	-0.3803	-0.0131	-0.0023	-0.0596
	$n = 300$	0.0017	-0.0282	-0.3962	-0.0259	0.0063	-0.0531
	$n = 450$	-0.0044	-0.0413	-0.3973	-0.0168	-0.0072	-0.0577
	$n = 600$	-0.0021	-0.0276	-0.3987	-0.0144	-0.0130	-0.0547
$\widehat{\beta}_2$	$n = 50$	0.0008	0.0079	0.0024	0.0068	-0.0019	0.0085
	$n = 150$	0.0014	0.0041	-0.0005	0.0023	0.0004	0.0043
	$n = 300$	<0.0001	0.0020	-0.0005	0.0018	-0.0008	0.0034
	$n = 450$	0.0003	0.0027	-0.0001	0.0013	0.0003	0.0039
	$n = 600$	0.0002	0.0019	0.0004	0.0017	0.0005	0.0034
MSE							
		SN		**ST**		**SCN**	
		`lqr()`	SMSN-CR	`lqr()`	SMSN-CR	`lqr()`	SMSN-CR
$\widehat{\beta}_1$	$n = 50$	0.0863	0.1678	0.3424	0.3555	0.2499	0.2120
	$n = 150$	0.0265	0.0469	0.1924	0.0919	0.0734	0.0717
	$n = 300$	0.0148	0.0264	0.1801	0.0428	0.0348	0.0357
	$n = 450$	0.0101	0.0173	0.1735	0.0280	0.0236	0.0251
	$n = 600$	0.0068	0.0143	0.1698	0.0202	0.0180	0.0184
$\widehat{\beta}_2$	$n = 50$	0.0006	0.0009	0.0011	0.0015	0.0017	0.0011
	$n = 150$	0.0002	0.0002	0.0003	0.0004	0.0005	0.0003
	$n = 300$	0.0001	0.0001	0.0002	0.0002	0.0002	0.0002
	$n = 450$	0.0001	0.0001	0.0001	0.0001	0.0002	0.0001
	$n = 600$	<0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

References

[1] C.E. Galarza, L. Benites, and V.H. Lachos, *lqr: Robust Linear Quantile Regression* (2016). Available at https://cran.r-project.org/web/packages/lqr, R package v. 1.5.

Table D.3: Simulation Study. Bias and MSE for the parameters β_1 and β_2, obtained by using our proposed SAEM algorithm and the *lqr*(·) package, with $p = 35\%$.

		Bias					
		SN		**ST**		**SCN**	
		`lqr()`	SMSN-CR	`lqr()`	SMSN-CR	`lqr()`	SMSN-CR
$\widehat{\beta}_1$	$n = 50$	0.0136	-0.1228	-0.3600	-0.0404	-0.0395	-0.2112
	$n = 150$	0.0050	-0.0684	-0.3907	-0.0215	-0.0178	-0.1653
	$n = 300$	-0.0090	-0.0898	-0.3886	-0.0160	0.0091	-0.1290
	$n = 450$	0.0026	-0.0621	-0.4052	-0.0080	0.0008	-0.1271
	$n = 600$	-0.0024	-0.0646	-0.4010	0.0026	-0.0043	-0.1243
$\widehat{\beta}_2$	$n = 50$	-0.0012	0.0077	-0.0011	0.0036	0.0037	0.0139
	$n = 150$	-0.0006	0.0043	0.0003	0.0027	0.0012	0.0108
	$n = 300$	0.0006	0.0059	-0.0007	0.0014	-0.0005	0.0086
	$n = 450$	0.0001	0.0043	0.0005	0.0012	0.0005	0.0089
	$n = 600$	0.0001	0.0041	0.0001	0.0005	0.0001	0.0082
		MSE					
		SN		**ST**		**SCN**	
		`lqr()`	SMSN-CR	`lqr()`	SMSN-CR	`lqr()`	SMSN-CR
$\widehat{\beta}_1$	$n = 50$	0.1024	0.9358	0.2765	0.8298	0.2141	0.7255
	$n = 150$	0.0316	0.1110	0.2031	0.2161	0.0669	0.1729
	$n = 300$	0.0145	0.0609	0.1745	0.0910	0.0353	0.0839
	$n = 450$	0.0102	0.0354	0.1795	0.0578	0.0281	0.0634
	$n = 600$	0.0076	0.0286	0.1731	0.0511	0.0173	0.0462
$\widehat{\beta}_2$	$n = 50$	0.0007	0.0041	0.0011	0.0034	0.0015	0.0032
	$n = 150$	0.0002	0.0005	0.0003	0.0009	0.0005	0.0008
	$n = 300$	0.0001	0.0003	0.0001	0.0004	0.0002	0.0004
	$n = 450$	0.0001	0.0002	0.0001	0.0002	0.0002	0.0003
	$n = 600$	<0.0001	0.0001	0.0001	0.0002	0.0001	0.0002