Supplemental Materials for "Partially Linear Structure Selection in Cox Models with Varying Coefficients"

by Heng Lian, Peng Lai and Hua Liang

Web Appendix A

Here we give the proofs of the main results. For notational simplicity, we take $K_j \equiv K$, and let C denote a generic constant that might assume different values at different places.

Proof of Theorem 1.

Let $\mathbf{a}_0 = (\mathbf{a}_{01}^T, \ldots, \mathbf{a}_{0p}^T)^T$ be a pK dimensional vector that satisfies $\|\alpha_{0j} - \mathbf{a}_{0j}^T\mathbf{B}\|_\infty = O(K^{-d}), 1 \leqslant j \leqslant s$ (such approximation rates are possible due to our smoothness assumption (C2) and well-know approximation properties of B-splines) and $\mathbf{a}_{0j} = \mathbf{0}, j > s$. Let $r_n = D(\sqrt{K/n} + K^{-d} + \lambda_1 + \lambda_2)$ for some constant $D > 0$ and $\mathbf{u} \in R^{pK}$ with $\|\mathbf{u}\| = 1$. We have

$$\begin{aligned}
\mathcal{L}(\mathbf{a}_0 + r_n\mathbf{u}) - \mathcal{L}(\mathbf{a}_0) &= \mathcal{Q}(\mathbf{a}_0 + r_n\mathbf{u}) - \mathcal{Q}(\mathbf{a}_0) \\
&\quad -n\sum_j p_{\lambda_1}(\|\mathbf{a}_{0j} + r_n\mathbf{u}_j\|) + n\sum_j p_{\lambda_1}(\|\mathbf{a}_{0j}\|) \\
&\quad -n\sum_j p_{\lambda_2}(\|\mathbf{a}_{0j} + r_n\mathbf{u}_j\|_c) + n\sum_j p_{\lambda_2}(\|\mathbf{a}_{0j}\|_c) \\
&\leqslant \mathcal{Q}(\mathbf{a}_0 + r_n\mathbf{u}) - \mathcal{Q}(\mathbf{a}_0) \\
&\quad +Cn(\lambda_1 + \lambda_2)\sum_j r_n||\mathbf{u}_j||, \\
&= \mathcal{Q}(\mathbf{a}_0 + r_n\mathbf{u}) - \mathcal{Q}(\mathbf{a}_0) + Cpn(\lambda_1 + \lambda_2)r_n, \qquad \text{(A.1)}
\end{aligned}$$

where in the inequality above we used $|p_\lambda(|s|) - p_\lambda(|t|)| \leqslant \lambda|s - t|$.

Using (5.3) and Lemma A.6 in Huang (1999), we have

$$\begin{aligned}\mathcal{Q}(\mathbf{a}_0 + r_n\mathbf{u}) - \mathcal{Q}(\mathbf{a}_0) &\leqslant n^{1/2}r_n(\sqrt{K} + \sqrt{\log(1/r_n)}) + E\mathcal{Q}(\mathbf{a}_0 + r_n\mathbf{u}) - E\mathcal{Q}(\mathbf{a}_0)\\ &\leqslant n^{1/2}r_n(\sqrt{K} + \sqrt{\log(1/r_n)}) - Cnr_n^2 + O_p(K + nK^{-2d}).\end{aligned}$$

Thus (A.1) is smaller than zero if D (in the definition of r_n) is chosen to be large enough. This shows the existence of a local maximizer $\hat{\mathbf{a}}$ with $\|\hat{\mathbf{a}} - \mathbf{a}_0\| = O_p(r_n) = O_p(\sqrt{K/n} + K^{-d} + \lambda)$.

The above convergence rate can be further improved to $||\widehat{\mathbf{a}} - \mathbf{a}_0||^2 = O_p(K/n + 1/K^{2d})$ as follows. First note that since the model is fixed as $n \to \infty$, we can find a constant $C > 0$ such that $\|\mathbf{a}_{0j}\| > C$ when $j \leqslant s$ and $\|\mathbf{a}_{0j}\|_c > C$ when $j \leqslant p_1$. Since $||\widehat{\mathbf{a}} - \mathbf{a}_0||^2 = o_p(1)$ by the convergence rates proved above and that $\lambda_1, \lambda_2 = o(1)$, we have $P\left(p_{\lambda_1}(\|\mathbf{a}_{0j}\|) = p_{\lambda_1}(\|\widehat{\mathbf{a}}_j\|)\right) \to 1$ if $j \leqslant s$. Similarly $P\left(p_{\lambda_2}(\|\mathbf{a}_{0j}\|_c) = p_{\lambda_2}(\|\widehat{\mathbf{a}}_j\|_c)\right) \to 1$ if $j \leqslant p_1$. These facts imply that

$$n\sum_{j=1}^{p} p_{\lambda_1}(\|\widehat{\mathbf{a}}_j\|) - n\sum_{j=1}^{p} p_{\lambda_1}(\|\mathbf{a}_{0j}\|) \geqslant 0,$$

and

$$n\sum_{j=1}^{p} p_{\lambda_2}(\|\widehat{\mathbf{a}}_j\|_c) - n\sum_{j=1}^{p} p_{\lambda_2}(\|\mathbf{a}_{0j}\|_c) \geqslant 0,$$

with probability tending to 1. Removing the regularizing terms in (A.1) and using the same reasoning as before, the rates are improved to $||\widehat{\mathbf{a}} - \mathbf{a}_0||^2 = O_p(K/n + 1/K^{2d})$.

The rates of convergence for $||\widehat{\mathbf{a}}_j - \mathbf{a}_{0j}||^2$ immediately imply the rates for $||\widehat{\alpha}_j(w) - \alpha_{0j}(w)||^2$ by the property (de Boor, 2001); that is,

$$C_1||\widehat{\mathbf{a}}_j^T\mathbf{B}_j - \mathbf{a}_{0j}^T\mathbf{B}_j||^2 \leqslant ||\widehat{\mathbf{a}}_j - \mathbf{a}_{0j}||^2 \leqslant C_2||\widehat{\mathbf{a}}_j^T\mathbf{B}_j - \mathbf{a}_{0j}^T\mathbf{B}_j||^2$$

for some constants $C_1, C_2 > 0$.

Proof of Theorem 2.

We only show part (b) concerning constansistency as an illustration since proof for part (a) is similar. Suppose for some $p_1 < j \leqslant s$, $\widehat{\mathbf{a}}_j^T\mathbf{B}$ does not represent a constant. Define $\widehat{\mathbf{a}}^*$ to be same as $\widehat{\mathbf{a}}$ except that $\widehat{\mathbf{a}}_j$ is replaced by its projection onto the subspace $\mathcal{M}$, that is $\widehat{\mathbf{a}}_j^* = \bar{\widehat{\mathbf{a}}}_j\mathbf{1}_K$, where $\bar{\widehat{\mathbf{a}}}_j = (\sum_k \widehat{\mathbf{a}}_{jk})/K$. Noting that $\|\widehat{\mathbf{a}}_j^*\|_c = 0$, we have

$$\begin{aligned} 0 &\leqslant \mathcal{L}(\widehat{\mathbf{a}}) - \mathcal{L}(\widehat{\mathbf{a}}^*) \\ &= \mathcal{Q}(\widehat{\mathbf{a}}) - \mathcal{Q}(\widehat{\mathbf{a}}^*) - np_{\lambda_1}(\|\widehat{\mathbf{a}}_j\|) + np_{\lambda_1}(\|\widehat{\mathbf{a}}_j^*\|) - np_{\lambda_2}(\|\widehat{\mathbf{a}}_j\|_c). \end{aligned}$$

As in the proof of Theorem 1, we have $P\left(p_{\lambda_1}(\|\widehat{\mathbf{a}}_j\|) \geqslant p_{\lambda_1}(\|\widehat{\mathbf{a}}_j^*\|)\right) \to 1$, and thus

$$0 \leqslant \mathcal{Q}(\widehat{\mathbf{a}}) - \mathcal{Q}(\widehat{\mathbf{a}}^*) - np_{\lambda_2}(\|\widehat{\mathbf{a}}_j\|_c).$$

Similar to the proof of Theorem 1, we have the bound

$$|\mathcal{Q}(\widehat{\mathbf{a}}) - \mathcal{Q}(\widehat{\mathbf{a}}^*)| = O_p(\sqrt{nK} + nK^{-d})\|\widehat{\mathbf{a}} - \widehat{\mathbf{a}}^*\| = O_p(\sqrt{nK} + nK^{-d})\|\widehat{\mathbf{a}}_j\|. \quad \text{(A.2)}$$

On the other hand, since $||\hat{\mathbf{a}}_j||_c = ||\hat{\mathbf{a}}_j - \mathbf{a}_{0j}||_c = O_p((K/n)^{1/2} + K^{-d}) = o(\lambda_2)$, we have

$$p_{\lambda_2}(||\hat{\mathbf{a}}_j||_c) = \lambda_2||\hat{\mathbf{a}}_j||_c, \text{with probability tending to 1,} \quad \text{(A.3)}$$

by the property of the SCAD penalty function. Noting that $||\hat{\mathbf{a}} - \hat{\mathbf{a}}^*|| = ||\hat{\mathbf{a}}_j - \hat{\mathbf{a}}_j^*|| = ||\hat{\mathbf{a}}_j||_c$ and combining (A.2) and (A.3), we have a contradiction if $||\hat{\mathbf{a}}_j||_c > 0$.

Proof of Theorem 3.

Given that the correct structure is selected (this happens with probability tending to one by Theorem 2), maximization of the penalization function is same as the maximization of the partial likelihood based on the correct structure. The asymptotic normality immediately follows from Huang (1999) since as we noted when using splines varying-coefficient models is almost identical to additive models in asymptotic theory.

References

de Boor, C. (2001). *A practical guide to splines.* Springer-Verlag, New York, rev. edition.

Huang, J. (1999). Efficient estimation of the partly linear additive Cox model. *Annals of statistics* **27,** 1536–1563.

Web Appendix B

Here we present more simulation results.

Example 1. This example is the same as the example presented in the main text except for that we set $n = 600$. The results are reported in Tables B.1-B.4 and Figure B.1. Figure B.2 plots some estimates of the coefficients for both $n = 300$ and $n = 600$.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Figure 1 about here.]

[Figure 2 about here.]

Example 2. In this example, we generate data from a model with three varying coefficients with $\alpha_{01}(w) = 10w(1-w), \alpha_{02}(w) = 2\sin(2\pi w), \alpha_{03}(w) = 2\sin(2\pi w)/(2 - \sin(2\pi w))$. We still have three constant coefficients $(\alpha_{03}, \alpha_{04}, \alpha_{05}) = (2, 1, -2)$ and $p = 8$. In this example, we only use $\nu = 0.04$ which results in roughly 20% censored observations. Other aspects of data generation are the same as before. Tables B.5 and B.6 show the variable selection results and the MSE for $n = 300$ respectively. Tables B.7 and B.8 are for $n = 600$. Figure B.3 shows the prediction errors. Coefficients estimates for some generated datasets are shown in Figure B.4.

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

Example 3. (Higher dimension) In this example, we use the setup of Example 2 with $\nu = 0.04$ and $\rho = 0.1$. Different from the previous example, we increase the dimension of the full model to $p = 20$ (with 3 varying coefficients and 3 constant coefficients). The results are shown in Tables B.9 and B.10, and Figure B.5. The results are still reasonably good although the computation is much slower than $p = 8$.

[Table 9 about here.]

[Table 10 about here.]

[Figure 5 about here.]

Example 4. (Effects of knots) In this example, we use the same setup as in Examples 2 and 3, with $\nu = 0.04$, $\rho = 0.1$, $n = 300$ and $p = 8, 20$. We consider different choice of K to study the effect of K on estimation. In particular, we consider $K = 4, 5, 6, 7, 8$ (note $K = 4$ is the smallest possible number of basis functions when cubic spline is used). The results are shown in Tables B.11, B.12 and Figure B.6 (here we only show the doubly penalized estimator, not estimator with a single penalty). To facilitate comparison, in the figure we restrict the y-range to $[0, 0.05]$ in all panels. From Figure B.6, we see that the results are similar for $K = 4, 5, 6, 7$ but much worse for $K = 8$, which is due to overfitting. Although not shown here, we also consider $n = 600$. For this larger sample size, all results are similar for K from 4 to 8. We did not try larger K due to computational time constraint, although one naturally believes if K is too large overfitting will occur and the performance will become worse.

[Table 11 about here.]

[Table 12 about here.]

[Figure 6 about here.]

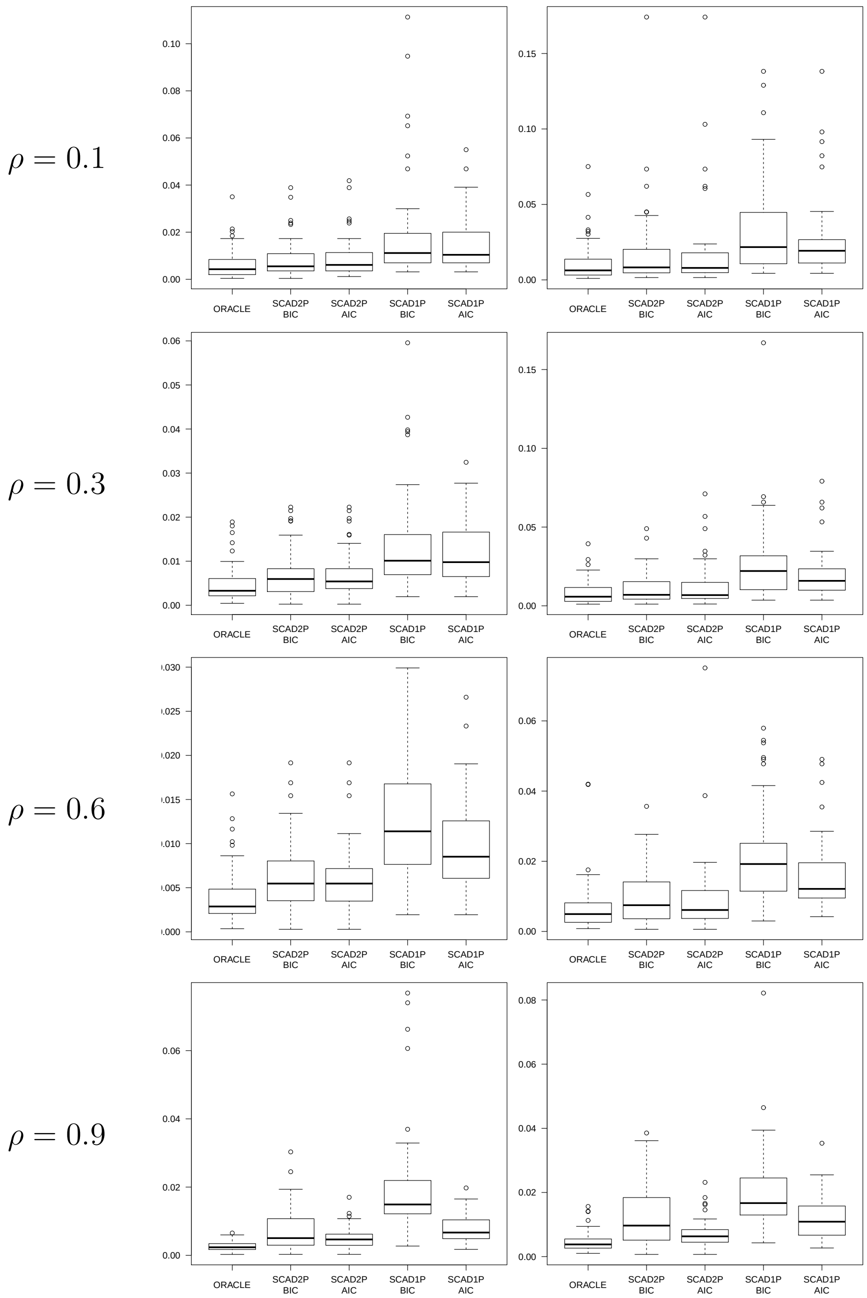

Figure B.1. For Example 1. Summarizing prediction errors for the simulation example with $n = 600$. Left column: $\nu = 0.04$ (20% censoring); Right column: $\nu = 0.15$ (50% censoring).

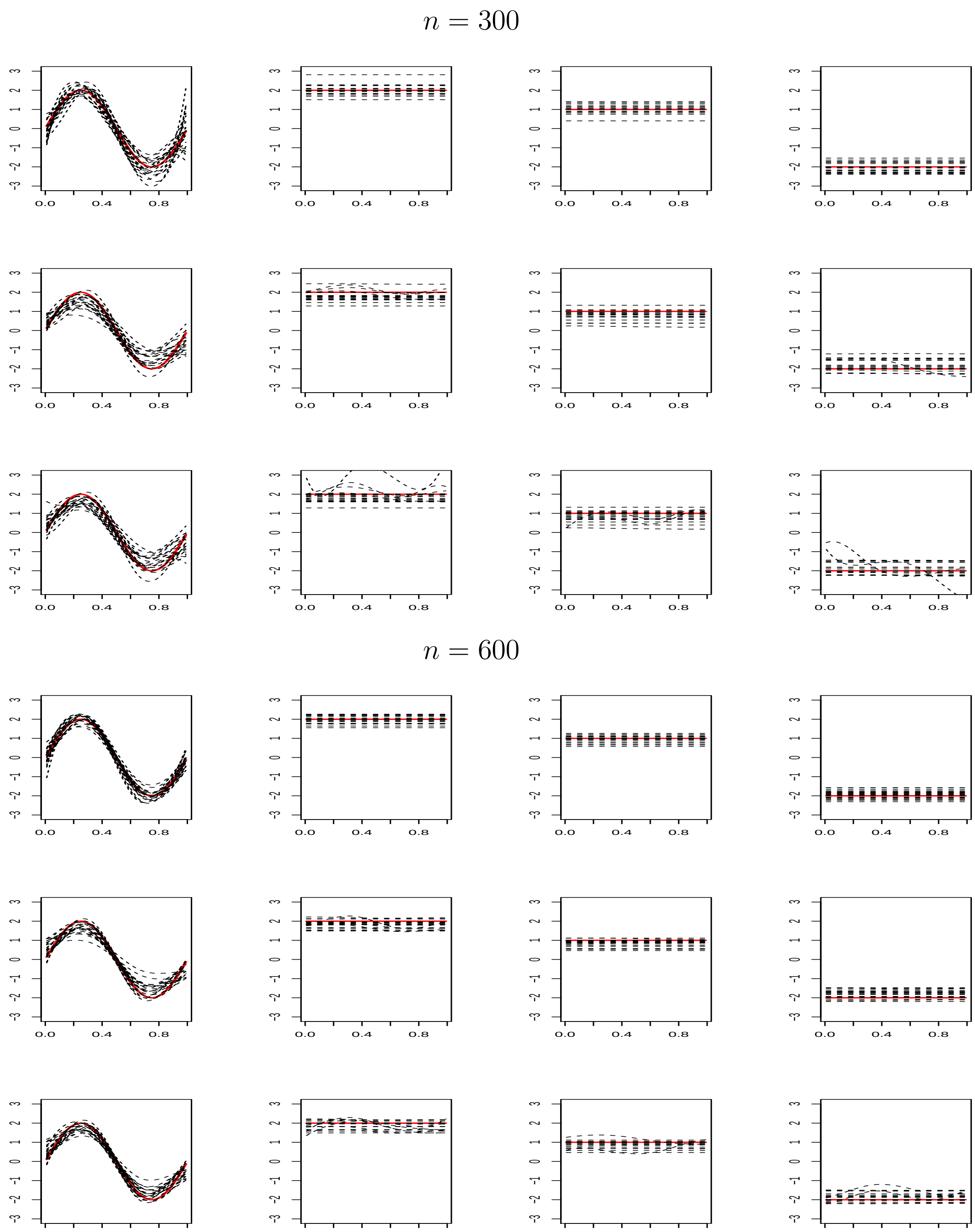

Figure B.2. Plots showing randomly picked 20 estimates for the four nonzero coefficients in Example 1 with $\rho = 0.1$ and $\nu = 0.04$ (25% censoring). The first three rows show the results of the oracle estimator, doubly penalized estimator with BIC, and doubly penalized estimator with AIC respectively, for $n = 300$. The last three rows show the results for $n = 600$. The red solid curves represent the true coefficient values.

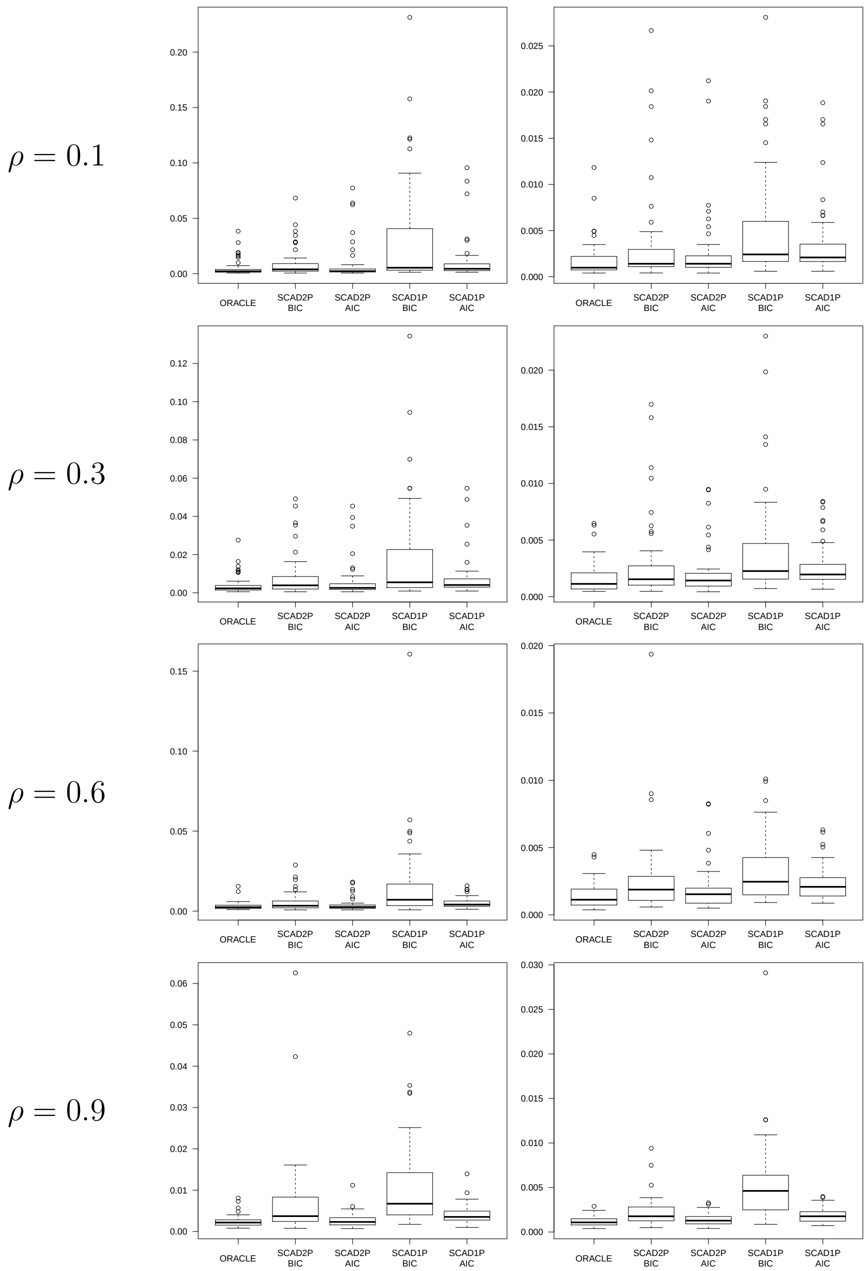

Figure B.3. For Example 2. Summarizing prediction errors for the simulation example. Left column: $n = 300$; right column: $n = 600$.

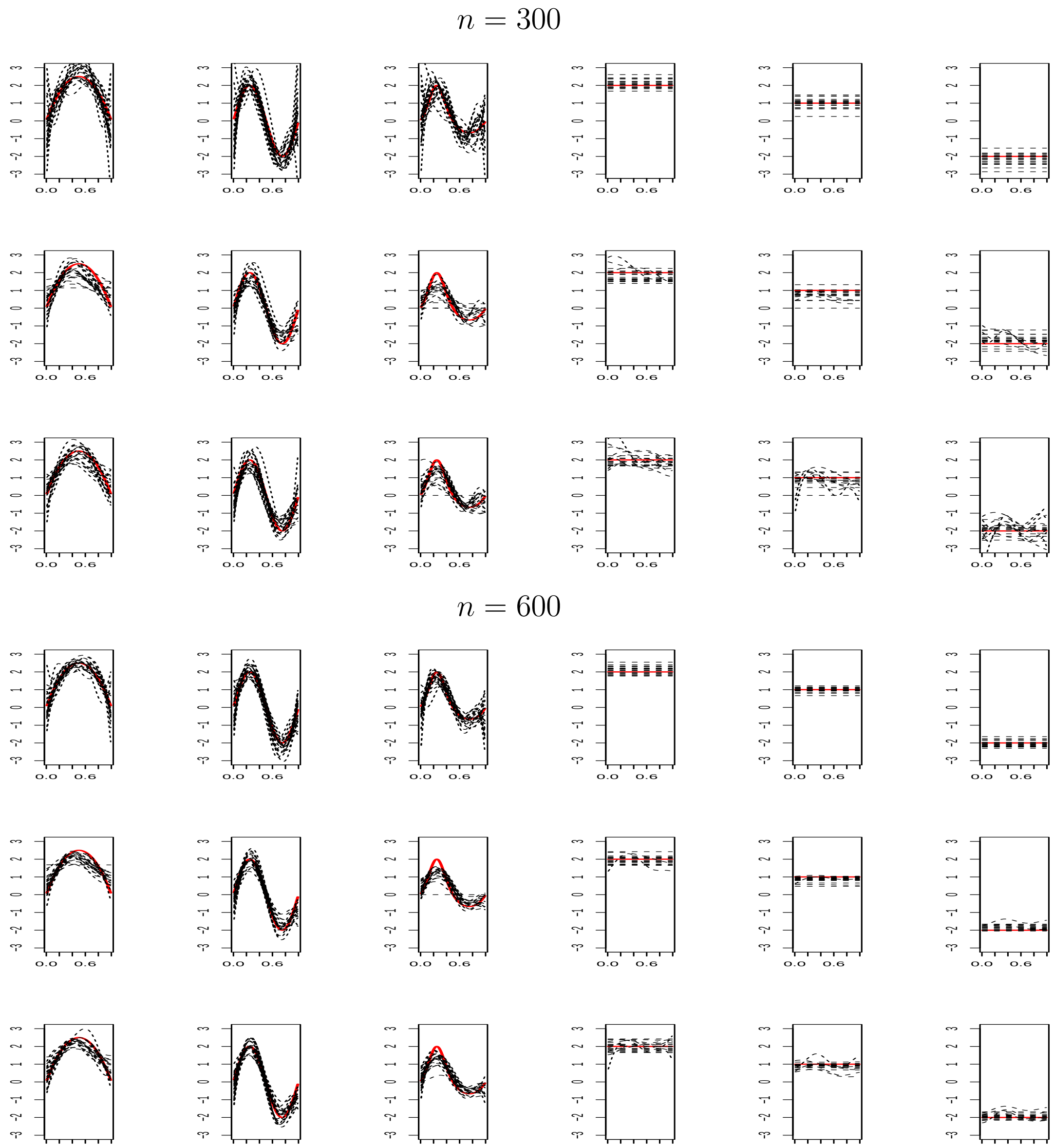

Figure B.4. For Example 2. Plots showing estimates for the six nonzero coefficients in Example 2 with $\rho = 0.1$ and $\nu = 0.04$ (20% censoring), with 20 randomly generated datasets. The first three rows show the results of the oracle estimator, doubly penalized estimator with BIC, and doubly penalized estimator with AIC respectively, for $n = 300$. The last three rows show the results for $n = 600$.

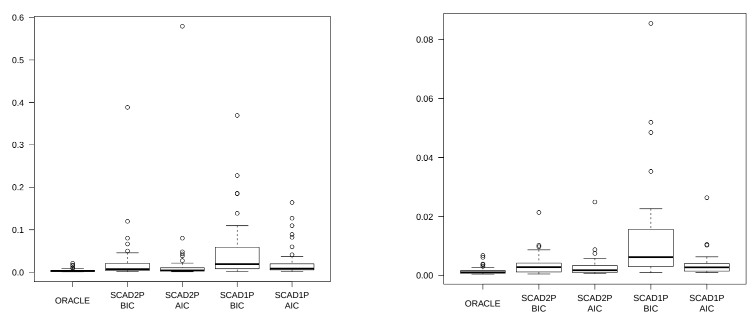

Figure B.5. For Example 3. Summarizing prediction errors for the simulation example. Left panel: $n = 300$; right panel: $n = 600$.

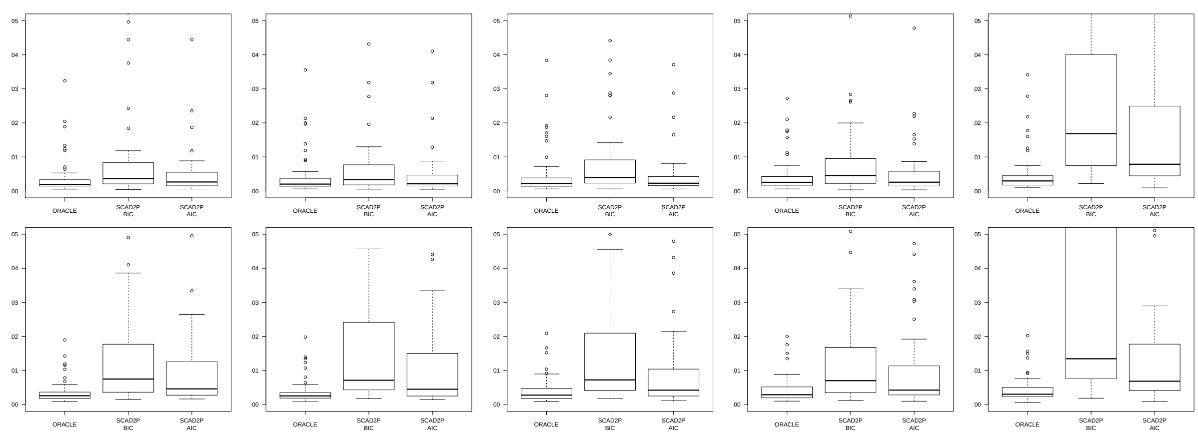

Figure B.6. For Example 4. Summarizing prediction errors for the simulation example as K changes from 4 to 8. The first row corresponds to $p = 8$ and the second row corresponds to $p = 20$.

Table B.1

For Example 1. Model selection results for SCAD estimators with either one penalty or two penalties, using AIC and BIC for tuning parameter selection, with $n = 600$ and $\nu = 0.04$ (25% censoring).

n	Method	#nonpar	#correct nonpar	#con	#correct con
0.1	SCAD2P.BIC	1.12(0.38)	1(0)	3.03(0.51)	2.88(0.38)
	SCAD2P.AIC	1.53(0.78)	1(0)	2.98(0.93)	2.56(0.70)
	SCAD1P.BIC	3.97(0.28)	3.94(0.24)	N/A	N/A
	SCAD1P.AIC	4.46(0.81)	4(0)	N/A	N/A
0.3	SCAD2P.BIC	1.16(0.42)	1(0)	3.02(0.55)	2.84(0.42)
	SCAD2P.AIC	1.49(0.67)	1(0)	3(0.88)	2.58(0.60)
	SCAD1P.BIC	3.96(0.45)	3.87(0.32)	N/A	N/A
	SCAD1P.AIC	4.61(0.94)	4(0)	N/A	N/A
0.6	SCAD2P.BIC	1.20(0.45)	1(0)	3.02(0.65)	2.8(0.45)
	SCAD2P.AIC	1.43(0.64)	1(0)	3(0.94)	2.6(0.63)
	SCAD1P.BIC	3.58(0.53)	3.56(0.50)	N/A	N/A
	SCAD1P.AIC	4.54(0.97)	3.95(0.24)	N/A	N/A
0.9	SCAD2P.BIC	1.4(0.60)	0.86(0.35)	2.29(1.39)	1.83(1.11)
	SCAD2P.AIC	1.80(0.75)	0.98(0.14)	2.62(1.21)	1.97(0.91)
	SCAD1P.BIC	2(0.85)	1.98(0.82)	N/A	N/A
	SCAD1P.AIC	4.06(1.38)	3.23(0.79)	N/A	N/A

Table B.2

For Example 1. Model selection results for SCAD estimators with either one penalty or two penalties, using AIC and BIC for tuning parameter selection, with $n = 600$ and $\nu = 0.15$ (50% censoring).

n	Method	#nonpar	#correct nonpar	#con	#correct con
0.1	SCAD2P.BIC	1.14(0.45)	1(0)	2.98(0.55)	2.83(0.46)
	SCAD2P.AIC	1.41(0.75)	1(0)	2.94(0.91)	2.64(0.72)
	SCAD1P.BIC	3.95(0.49)	3.85(0.35)	*N/A*	*N/A*
	SCAD1P.AIC	4.72(1.07)	3.98(0.14)	*N/A*	*N/A*
0.3	SCAD2P.BIC	1.21(0.45)	1(0)	3.02(0.58)	2.83(0.43)
	SCAD2P.AIC	1.57(0.81)	1(0)	2.88(1.06)	2.48(0.78)
	SCAD1P.BIC	3.76(0.59)	3.67(0.51)	*N/A*	*N/A*
	SCAD1P.AIC	4.70(1.04)	3.94(0.24)	*N/A*	*N/A*
0.6	SCAD2P.BIC	1.34(0.55)	1(0)	2.76(0.87)	2.56(0.73)
	SCAD2P.AIC	1.60(0.80)	1(0)	2.81(1.18)	2.34(0.89)
	SCAD1P.BIC	3.40(0.72)	3.31(0.64)	*N/A*	*N/A*
	SCAD1P.AIC	4.67(1.08)	3.83(0.40)	*N/A*	*N/A*
0.9	SCAD2P.BIC	1.31(0.54)	0.72(0.45)	1.66(1.56)	1.23(1.15)
	SCAD2P.AIC	1.83(0.82)	0.96(0.19)	2.17(1.40)	1.59(1.03)
	SCAD1P.BIC	1.68(0.71)	1.66(0.68)	*N/A*	*N/A*
	SCAD1P.AIC	3.82(1.76)	2.94(1.06)	*N/A*	*N/A*

Table B.3

For Example 1. MSE for five estimated coefficients, with $n = 600$ and $\nu = 0.04$ (25% censoring).

ρ	Method	α_1	α_2	α_3	α_4	α_5
0.1	ORACLE	0.066(0.052)	0.039(0.043)	0.028(0.037)	0.041(0.049)	N/A
	SCAD2P.BIC	0.135(0.108)	0.066(0.075)	0.044(0.063)	0.064(0.072)	0(0)
	SCAD2P.AIC	0.108(0.072)	0.062(0.064)	0.046(0.065)	0.062(0.071)	0.002(0.011)
	SCAD1P.BIC	0.195(0.126)	0.229(0.173)	0.239(0.229)	0.282(0.226)	0(0)
	SCAD1P.AIC	0.173(0.089)	0.187(0.109)	0.175(0.107)	0.207(0.117)	0.011(0.040)
0.3	ORACLE	0.067(0.056)	0.044(0.051)	0.037(0.041)	0.046(0.053)	N/A
	SCAD2P.BIC	0.131(0.106)	0.074(0.086)	0.058(0.077)	0.070(0.078)	0(0)
	SCAD2P.AIC	0.118(0.104)	0.070(0.075)	0.058(0.071)	0.073(0.074)	0.003(0.014)
	SCAD1P.BIC	0.228(0.143)	0.241(0.168)	0.307(0.285)	0.391(0.341)	0(0)
	SCAD1P.AIC	0.189(0.095)	0.224(0.129)	0.202(0.128)	0.233(0.131)	0.013(0.046)
0.6	ORACLE	0.081(0.069)	0.071(0.081)	0.063(0.068)	0.062(0.070)	N/A
	SCAD2P.BIC	0.165(0.173)	0.118(0.138)	0.118(0.160)	0.127(0.134)	0.009(0.032)
	SCAD2P.AIC	0.148(0.146)	0.110(0.125)	0.108(0.142)	0.124(0.129)	0.011(0.030)
	SCAD1P.BIC	0.381(0.231)	0.393(0.241)	0.599(0.388)	0.940(0.720)	0(0)
	SCAD1P.AIC	0.256(0.145)	0.352(0.213)	0.347(0.248)	0.403(0.296)	0.021(0.062)
0.9	ORACLE	0.186(0.232)	0.274(0.320)	0.235(0.276)	0.177(0.214)	N/A
	SCAD2P.BIC	0.584(0.669)	0.676(0.853)	0.588(0.433)	1.240(1.514)	0.039(0.124)
	SCAD2P.AIC	0.401(0.460)	0.442(0.487)	0.489(0.427)	0.707(0.885)	0.048(0.124)
	SCAD1P.BIC	1.233(0.695)	2.082(0.936)	0.959(0.148)	3.521(1.052)	0.001(0.002)
	SCAD1P.AIC	0.795(0.396)	1.247(0.889)	0.932(0.524)	1.820(1.172)	0.056(0.161)

Table B.4

For Example 1. MSE for five estimated coefficients, with $n = 600$ and $\nu = 0.15$ (50% censoring).

ρ	Method	α_1	α_2	α_3	α_4	α_5
0.1	ORACLE	0.113(0.085)	0.058(0.062)	0.051(0.081)	0.063(0.083)	N/A
	SCAD2P.BIC	0.181(0.142)	0.087(0.124)	0.090(0.168)	0.086(0.126)	0.006(0.031)
	SCAD2P.AIC	0.163(0.131)	0.082(0.108)	0.078(0.134)	0.082(0.116)	0.009(0.035)
	SCAD1P.BIC	0.294(0.216)	0.334(0.275)	0.364(0.306)	0.414(0.339)	0(0)
	SCAD1P.AIC	0.264(0.156)	0.275(0.182)	0.261(0.189)	0.322(0.233)	0.047(0.126)
0.3	ORACLE	0.111(0.085)	0.068(0.075)	0.059(0.094)	0.063(0.077)	N/A
	SCAD2P.BIC	0.199(0.165)	0.094(0.125)	0.098(0.170)	0.095(0.115)	0.008(0.029)
	SCAD2P.AIC	0.163(0.136)	0.101(0.123)	0.093(0.169)	0.101(0.133)	0.010(0.029)
	SCAD1P.BIC	0.385(0.312)	0.424(0.359)	0.491(0.366)	0.708(0.685)	0(0)
	SCAD1P.AIC	0.284(0.153)	0.339(0.215)	0.312(0.232)	0.377(0.258)	0.037(0.102)
0.6	ORACLE	0.128(0.112)	0.114(0.141)	0.115(0.166)	0.097(0.113)	N/A
	SCAD2P.BIC	0.248(0.244)	0.164(0.209)	0.209(0.316)	0.217(0.343)	0.010(0.035)
	SCAD2P.AIC	0.222(0.220)	0.167(0.214)	0.191(0.293)	0.202(0.325)	0.014(0.037)
	SCAD1P.BIC	0.658(0.373)	0.719(0.559)	0.757(0.335)	1.430(1.075)	0.001(0.010)
	SCAD1P.AIC	0.415(0.247)	0.523(0.353)	0.510(0.311)	0.626(0.442)	0.057(0.118)
0.9	ORACLE	0.288(0.304)	0.483(0.582)	0.526(0.696)	0.337(0.412)	N/A
	SCAD2P.BIC	0.910(0.850)	1.113(1.230)	0.744(0.377)	2.283(1.737)	0.068(0.231)
	SCAD2P.AIC	0.635(0.646)	0.779(0.870)	0.627(0.431)	1.270(1.453)	0.085(0.216)
	SCAD1P.BIC	1.280(0.678)	2.450(0.841)	0.968(0.109)	3.890(0.546)	0.001(0.004)
	SCAD1P.AIC	1.060(0.602)	1.762(1.046)	0.980(0.434)	2.313(1.387)	0.137(0.415)

Table B.5

For Example 2. Model selection results for SCAD estimators with either one penalty or two penalties, using AIC and BIC for tuning parameter selection, with $n = 300$ and $\nu = 0.04$ (20% censoring).

ρ	Method	#nonpar	#correct nonpar	#con	#correct con
0.1	SCAD2P.BIC	3.08(0.96)	2.66(0.62)	2.88(1.02)	2.56(0.70)
	SCAD2P.AIC	4.44(1.21)	2.98(0.14)	1.98(1.21)	1.71(1.04)
	SCAD1P.BIC	5.68(0.89)	5.63(0.78)	*N/A*	*N/A*
	SCAD1P.AIC	6.90(0.93)	6.00(0)	*N/A*	*N/A*
0.3	SCAD2P.BIC	2.84(0.97)	2.52(0.73)	3.20(0.94)	2.66(0.55)
	SCAD2P.AIC	4.30(1.25)	2.95(0.19)	2.20(1.29)	1.84(1.09)
	SCAD1P.BIC	5.53(0.93)	5.46(0.83)	*N/A*	*N/A*
	SCAD1P.AIC	6.97(0.85)	5.99(0.14)	*N/A*	*N/A*
0.6	SCAD2P.BIC	2.78(0.99)	2.44(0.73)	3.24(1.08)	2.56(0.64)
	SCAD2P.AIC	4.02(1.25)	2.76(0.47)	2.37(1.38)	1.87(1.07)
	SCAD1P.BIC	4.64(1.37)	4.60(1.34)	*N/A*	*N/A*
	SCAD1P.AIC	6.85(0.86)	5.94(0.24)	*N/A*	*N/A*
0.9	SCAD2P.BIC	2.32(0.68)	1.92(0.69)	1.52(1.47)	1.22(1.15)
	SCAD2P.AIC	3.20(1.01)	2.40(0.70)	2.26(1.4)	1.72(1.05)
	SCAD1P.BIC	2.84(0.91)	2.83(0.91)	*N/A*	*N/A*
	SCAD1P.AIC	5.65(1.89)	4.85(1.26)	*N/A*	*N/A*

Table B.6

For Example 2. MSE for seven estimated coefficients, with $n = 300$ and $\nu = 0.04$ (20% censoring).

ρ	Method	α_1	α_2	α_3	α_4	α_5	α_6	α_7
0.1	ORACLE	0.327(0.244)	0.381(0.313)	0.358(0.219)	0.077(0.102)	0.063(0.101)	0.119(0.157)	*N/A*
	SCAD2P.BIC	0.331(0.211)	0.254(0.180)	0.357(0.309)	0.149(0.186)	0.114(0.167)	0.158(0.206)	0.002(0.013)
	SCAD2P.AIC	0.209(0.148)	0.230(0.171)	0.229(0.163)	0.129(0.183)	0.105(0.159)	0.166(0.161)	0.030(0.099)
	SCAD1P.BIC	0.468(0.359)	0.405(0.355)	0.377(0.306)	0.507(0.404)	0.482(0.334)	0.616(0.468)	0.077(0.383)
	SCAD1P.AIC	0.443(0.327)	0.472(0.356)	0.378(0.277)	0.438(0.331)	0.429(0.343)	0.592(0.448)	0.270(0.458)
0.3	ORACLE	0.377(0.281)	0.500(0.384)	0.430(0.256)	0.090(0.127)	0.076(0.115)	0.123(0.162)	*N/A*
	SCAD2P.BIC	0.378(0.258)	0.309(0.306)	0.387(0.321)	0.133(0.166)	0.124(0.185)	0.159(0.211)	0.004(0.018)
	SCAD2P.AIC	0.225(0.167)	0.248(0.191)	0.252(0.168)	0.126(0.168)	0.113(0.166)	0.177(0.173)	0.033(0.108)
	SCAD1P.BIC	0.501(0.363)	0.471(0.369)	0.425(0.319)	0.489(0.375)	0.566(0.363)	0.795(0.669)	0.036(0.256)
	SCAD1P.AIC	0.491(0.355)	0.534(0.358)	0.448(0.307)	0.528(0.421)	0.525(0.428)	0.634(0.473)	0.262(0.397)
0.6	ORACLE	0.577(0.429)	0.918(0.656)	0.663(0.385)	0.152(0.226)	0.143(0.242)	0.152(0.195)	*N/A*
	SCAD2P.BIC	0.424(0.237)	0.425(0.374)	0.441(0.329)	0.165(0.255)	0.249(0.317)	0.292(0.346)	0.019(0.054)
	SCAD2P.AIC	0.317(0.206)	0.391(0.334)	0.360(0.266)	0.201(0.276)	0.194(0.263)	0.262(0.294)	0.044(0.141)
	SCAD1P.BIC	0.715(0.536)	0.853(0.660)	0.569(0.319)	0.869(0.773)	0.855(0.355)	1.670(1.343)	0(0)
	SCAD1P.AIC	0.580(0.431)	0.803(0.563)	0.614(0.394)	0.842(0.592)	0.827(0.660)	0.904(0.690)	0.325(0.533)
0.9	ORACLE	2.046(1.613)	4.170(3.310)	2.633(1.440)	0.575(0.815)	0.642(1.070)	0.408(0.525)	*N/A*
	SCAD2P.BIC	1.223(0.712)	1.424(0.699)	0.758(0.383)	1.435(1.374)	0.874(0.366)	2.383(1.697)	0.052(0.259)
	SCAD2P.AIC	0.827(0.586)	1.035(0.681)	0.750(0.370)	0.804(0.912)	0.702(0.559)	1.300(1.313)	0.130(0.427)
	SCAD1P.BIC	1.435(0.535)	1.693(0.592)	0.789(0.371)	2.346(1.044)	0.972(0.119)	3.733(0.748)	0(0)
	SCAD1P.AIC	1.329(0.863)	1.853(1.135)	1.627(1.313)	2.335(1.590)	1.794(1.776)	2.800(1.596)	0.516(1.074)

Table B.7

For Example 2. Model selection results for SCAD estimators with either one penalty or two penalties, using AIC and BIC for tuning parameter selection, with $n = 600$ and $\nu = 0.04$ (20% censoring).

ρ	Method	#nonpar	#correct nonpar	#con	#correct con
0.1	SCAD2P.BIC	3.11(0.50)	2.87(0.32)	2.92(0.60)	2.77(0.41)
	SCAD2P.AIC	4.22(1.11)	3.00(0)	2.06(0.93)	2.00(0.96)
	SCAD1P.BIC	6.00(0.28)	5.95(0.19)	N/A	N/A
	SCAD1P.AIC	6.53(0.70)	6.00(0)	N/A	N/A
0.3	SCAD2P.BIC	3.02(0.47)	2.82(0.38)	3.04(0.63)	2.80(0.40)
	SCAD2P.AIC	3.96(0.94)	3.00(0)	2.26(0.92)	2.14(0.83)
	SCAD1P.BIC	5.95(0.46)	5.84(0.37)	N/A	N/A
	SCAD1P.AIC	6.62(0.69)	6.00(0)	N/A	N/A
0.6	SCAD2P.BIC	3.02(0.65)	2.72(0.49)	2.92(0.72)	2.66(0.51)
	SCAD2P.AIC	3.62(0.74)	2.86(0.35)	2.62(0.87)	2.31(0.68)
	SCAD1P.BIC	5.61(0.69)	5.58(0.60)	N/A	N/A
	SCAD1P.AIC	6.64(0.74)	5.97(0.14)	N/A	N/A
0.9	SCAD2P.BIC	2.52(0.83)	2.14(0.78)	2.41(1.43)	1.90(0.97)
	SCAD2P.AIC	3.20(1.01)	2.53(0.67)	2.60(1.09)	2.10(0.67)
	SCAD1P.BIC	3.40(1.07)	3.39(1.07)	N/A	N/A
	SCAD1P.AIC	6.28(1.25)	5.51(0.67)	N/A	N/A

Table B.8

For Example 2. MSE for seven estimated coefficients, with $n = 600$ and $\nu = 0.04$ (20% censoring).

ρ	Method	α_1	α_2	α_3	α_4	α_5	α_6	α_7
0.1	ORACLE	0.151(0.072)	0.162(0.107)	0.159(0.077)	0.046(0.067)	0.018(0.022)	0.029(0.038)	N/A
	SCAD2P.BIC	0.234(0.166)	0.148(0.093)	0.182(0.168)	0.054(0.061)	0.044(0.062)	0.048(0.068)	0(0)
	SCAD2P.AIC	0.148(0.091)	0.133(0.074)	0.142(0.089)	0.052(0.055)	0.037(0.051)	0.041(0.055)	0.006(0.023)
	SCAD1P.BIC	0.180(0.107)	0.176(0.130)	0.200(0.154)	0.196(0.150)	0.204(0.157)	0.191(0.134)	0(0)
	SCAD1P.AIC	0.190(0.111)	0.204(0.143)	0.186(0.113)	0.210(0.127)	0.170(0.090)	0.186(0.118)	0.040(0.087)
0.3	ORACLE	0.166(0.078)	0.208(0.136)	0.183(0.099)	0.053(0.079)	0.021(0.027)	0.029(0.040)	N/A
	SCAD2P.BIC	0.265(0.183)	0.181(0.125)	0.217(0.204)	0.058(0.065)	0.057(0.083)	0.067(0.114)	0.001(0.004)
	SCAD2P.AIC	0.175(0.114)	0.163(0.095)	0.163(0.102)	0.056(0.062)	0.041(0.053)	0.041(0.066)	0.003(0.014)
	SCAD1P.BIC	0.200(0.120)	0.199(0.144)	0.239(0.216)	0.226(0.154)	0.285(0.256)	0.247(0.205)	0(0)
	SCAD1P.AIC	0.192(0.116)	0.224(0.151)	0.216(0.142)	0.249(0.156)	0.195(0.109)	0.209(0.131)	0.043(0.089)
0.6	ORACLE	0.247(0.119)	0.388(0.250)	0.271(0.157)	0.089(0.114)	0.045(0.061)	0.034(0.049)	N/A
	SCAD2P.BIC	0.331(0.201)	0.272(0.141)	0.312(0.285)	0.095(0.107)	0.132(0.213)	0.091(0.123)	0.001(0.004)
	SCAD2P.AIC	0.251(0.156)	0.255(0.148)	0.253(0.222)	0.097(0.115)	0.084(0.115)	0.058(0.086)	0.004(0.016)
	SCAD1P.BIC	0.301(0.206)	0.315(0.203)	0.351(0.301)	0.325(0.164)	0.512(0.338)	0.460(0.356)	0.001(0.002)
	SCAD1P.AIC	0.271(0.174)	0.366(0.243)	0.328(0.263)	0.396(0.239)	0.328(0.187)	0.297(0.192)	0.074(0.156)
0.9	ORACLE	0.844(0.402)	1.675(1.033)	1.020(0.643)	0.336(0.437)	0.202(0.257)	0.109(0.138)	N/A
	SCAD2P.BIC	0.779(0.481)	0.911(0.569)	0.655(0.388)	0.545(0.639)	0.675(0.404)	1.150(1.438)	0.029(0.118)
	SCAD2P.AIC	0.608(0.339)	0.693(0.393)	0.557(0.345)	0.351(0.417)	0.459(0.382)	0.462(0.786)	0.042(0.144)
	SCAD1P.BIC	1.163(0.655)	1.410(0.602)	0.805(0.295)	1.640(0.834)	0.965(0.144)	3.030(1.397)	0.010(0.071)
	SCAD1P.AIC	0.701(0.428)	0.943(0.528)	0.858(0.782)	1.100(0.616)	0.853(0.342)	0.936(0.651)	0.204(0.393)

Table B.9

For Example 3. Model selection results for SCAD estimators with either one penalty or two penalties, using AIC and BIC for tuning parameter selection, with $p = 20$ and $\nu = 0.04$.

n	Method	#nonpar	#correct nonpar	#con	#correct con
$n = 300$	SCAD2P.BIC	2.93(1.10)	2.40(0.72)	3.42(1.54)	2.46(0.73)
	SCAD2P.AIC	4.68(1.42)	2.85(0.42)	3.55(2.18)	1.91(0.80)
	SCAD1P.BIC	5.51(0.99)	5.34(0.87)	N/A	N/A
	SCAD1P.AIC	9.06(5.00)	5.88(0.38)	N/A	N/A
$n = 600$	SCAD2P.BIC	3.22(0.61)	2.90(0.30)	3.27(1.07)	2.67(0.51)
	SCAD2P.AIC	4.32(1.15)	3(0)	3.30(1.78)	2.17(0.84)
	SCAD1P.BIC	5.91(0.50)	5.88(0.48)	N/A	N/A
	SCAD1P.AIC	7.38(1.41)	6(0)	N/A	N/A

Table B.10

For Example 3. MSE for seven estimated coefficients, with $p = 20$ and $\nu = 0.04$.

n	Method	α_1	α_2	α_3	α_4	α_5	α_6	α_7
$n = 300$	ORACLE	0.332(0.197)	0.367(0.190)	0.345(0.182)	0.089(0.142)	0.063(0.096)	0.073(0.087)	N/A
	SCAD2P.BIC	0.500(0.312)	0.340(0.213)	0.529(0.326)	0.224(0.189)	0.240(0.279)	0.307(0.315)	0.003(0.016)
	SCAD2P.AIC	0.291(0.178)	0.266(0.164)	0.324(0.244)	0.138(0.159)	0.138(0.193)	0.148(0.180)	0.021(0.052)
	SCAD1P.BIC	0.642(0.556)	0.489(0.396)	0.497(0.327)	0.656(0.477)	0.573(0.344)	0.750(0.660)	0(0)
	SCAD1P.AIC	0.524(0.379)	0.594(0.794)	0.514(0.446)	0.608(0.571)	0.518(0.515)	0.675(0.788)	0.162(0.434)
$n = 600$	ORACLE	0.146(0.089)	0.144(0.085)	0.168(0.087)	0.027(0.044)	0.029(0.040)	0.034(0.041)	N/A
	SCAD2P.BIC	0.246(0.187)	0.154(0.086)	0.18(0.104)	0.069(0.079)	0.064(0.073)	0.099(0.100)	0.0003(0.002)
	SCAD2P.AIC	0.149(0.097)	0.132(0.071)	0.134(0.079)	0.048(0.058)	0.057(0.054)	0.063(0.065)	0.004(0.015)
	SCAD1P.BIC	0.265(0.203)	0.215(0.105)	0.246(0.208)	0.245(0.164)	0.335(0.235)	0.327(0.314)	0(0)
	SCAD1P.AIC	0.171(0.088)	0.165(0.083)	0.169(0.086)	0.151(0.085)	0.192(0.094)	0.184(0.105)	0.002(0.008)

Table B.11

For Example 4. Model selection results for SCAD estimators with either one penalty or two penalties, using AIC and BIC for tuning parameter selection, with $n = 300$, $p = 8, 20$.

K	Method	#nonpar	#correct nonpar	#con	#correct con
			$p = 8$		
$K = 4$	SCAD2P.BIC	3.40(0.85)	2.82(0.38)	2.71(0.93)	2.47(0.76)
	SCAD2P.AIC	4.35(1.10)	2.94(0.24)	2.14(1.14)	1.80(0.99)
$K = 5$	SCAD2P.BIC	3.14(0.75)	2.78(0.46)	2.92(0.90)	2.62(0.63)
	SCAD2P.AIC	4.31(1.13)	2.96(0.19)	2.17(1.11)	1.84(0.91)
$K = 6$	SCAD2P.BIC	3.08(0.96)	2.66(0.62)	2.88(1.02)	2.56(0.70)
	SCAD2P.AIC	4.44(1.21)	2.98(0.14)	1.98(1.21)	1.71(1.04)
$K = 7$	SCAD2P.BIC	2.91(0.98)	2.53(0.73)	3.17(1.04)	2.60(0.60)
	SCAD2P.AIC	4.48(1.39)	2.95(0.19)	2.04(1.46)	1.67(1.06)
$K = 8$	SCAD2P.BIC	3.07(1.12)	2.38(0.75)	2.42(1.03)	2.16(0.79)
	SCAD2P.AIC	4.73(1.03)	2.92(0.27)	1.14(1.07)	1.13(1.02)
			$p = 20$		
$K = 4$	SCAD2P.BIC	3.72(0.94)	2.71(0.50)	2.87(1.36)	2.22(0.79)
	SCAD2P.AIC	5.42(1.51)	2.88(0.32)	3.17(1.65)	1.92(0.85)
$K = 5$	SCAD2P.BIC	3.11(1.04)	2.60(0.60)	3.13(1.36)	2.46(0.81)
	SCAD2P.AIC	4.96(1.52)	2.90(0.30)	3.59(2.29)	1.90(0.97)
$K = 6$	SCAD2P.BIC	2.93(1.10)	2.40(0.72)	3.42(1.54)	2.46(0.73)
	SCAD2P.AIC	4.68(1.42)	2.85(0.42)	3.55(2.18)	1.91(0.80)
$K = 7$	SCAD2P.BIC	3.16(1.11)	2.53(0.64)	3.42(1.83)	2.36(0.85)
	SCAD2P.AIC	4.63(1.47)	2.78(0.46)	3.35(2.66)	1.81(1.03)
$K = 8$	SCAD2P.BIC	3.06(1.13)	2.28(0.83)	2.34(1.15)	2.04(0.94)
	SCAD2P.AIC	4.71(1.36)	2.81(0.43)	1.42(1.16)	1.28(1.05)

Table B.12

For Example 4. MSE for seven estimated coefficients, with $n = 300$, $p = 8, 20$.

K	Method	α_1	α_2	α_3	α_4	α_5	α_6	α_7
				$p = 8$				
$K = 4$	ORACLE	0.193(0.152)	0.221(0.167)	0.279(0.173)	0.073(0.111)	0.059(0.106)	0.099(0.117)	N/A
	SCAD2P.BIC	0.287(0.180)	0.285(0.203)	0.344(0.265)	0.152(0.251)	0.103(0.165)	0.156(0.221)	0.007(0.028)
	SCAD2P.AIC	0.212(0.152)	0.209(0.155)	0.267(0.158)	0.107(0.149)	0.093(0.155)	0.130(0.120)	0.022(0.051)
$K = 5$	ORACLE	0.265(0.193)	0.305(0.252)	0.284(0.189)	0.070(0.099)	0.060(0.105)	0.099(0.128)	N/A
	SCAD2P.BIC	0.277(0.226)	0.279(0.287)	0.326(0.271)	0.131(0.180)	0.097(0.169)	0.132(0.168)	0.003(0.025)
	SCAD2P.AIC	0.194(0.138)	0.225(0.146)	0.235(0.165)	0.119(0.155)	0.086(0.156)	0.122(0.108)	0.031(0.094)
$K = 6$	ORACLE	0.327(0.244)	0.381(0.313)	0.358(0.219)	0.077(0.102)	0.063(0.101)	0.119(0.157)	N/A
	SCAD2P.BIC	0.331(0.211)	0.254(0.180)	0.357(0.309)	0.149(0.186)	0.114(0.167)	0.158(0.206)	0.002(0.013)
	SCAD2P.AIC	0.209(0.148)	0.230(0.171)	0.229(0.163)	0.129(0.183)	0.105(0.159)	0.166(0.161)	0.030(0.099)
$K = 7$	ORACLE	0.404(0.272)	0.479(0.358)	0.446(0.278)	0.090(0.115)	0.064(0.101)	0.123(0.151)	N/A
	SCAD2P.BIC	0.394(0.249)	0.291(0.284)	0.382(0.329)	0.158(0.205)	0.108(0.168)	0.176(0.219)	0.002(0.009)
	SCAD2P.AIC	0.225(0.157)	0.236(0.158)	0.226(0.158)	0.126(0.152)	0.106(0.164)	0.162(0.153)	0.034(0.086)
$K = 8$	ORACLE	0.493(0.296)	0.599(0.428)	0.528(0.294)	0.092(0.121)	0.065(0.101)	0.126(0.155)	N/A
	SCAD2P.BIC	0.618(0.343)	0.390(0.290)	0.497(0.349)	0.350(0.331)	0.369(0.310)	0.424(0.472)	0(0)
	SCAD2P.AIC	0.306(0.156)	0.280(0.204)	0.295(0.235)	0.257(0.244)	0.301(0.272)	0.307(0.293)	0.0002(0.002)
				$p = 20$				
$K = 4$	ORACLE	0.213(0.162)	0.264(0.185)	0.267(0.158)	0.080(0.119)	0.059(0.087)	0.064(0.083)	N/A
	SCAD2P.BIC	0.410(0.275)	0.324(0.251)	0.453(0.298)	0.226(0.247)	0.247(0.285)	0.255(0.294)	0.013(0.056)
	SCAD2P.AIC	0.279(0.155)	0.272(0.170)	0.335(0.212)	0.141(0.180)	0.116(0.136)	0.119(0.146)	0.032(0.093)
$K = 5$	ORACLE	0.282(0.177)	0.314(0.182)	0.265(0.144)	0.085(0.132)	0.060(0.088)	0.065(0.083)	N/A
	SCAD2P.BIC	0.452(0.295)	0.346(0.219)	0.481(0.325)	0.260(0.254)	0.242(0.274)	0.306(0.424)	0.006(0.035)
	SCAD2P.AIC	0.286(0.164)	0.293(0.193)	0.335(0.242)	0.149(0.170)	0.135(0.182)	0.136(0.154)	0.026(0.067)
$K = 6$	ORACLE	0.332(0.197)	0.367(0.190)	0.345(0.182)	0.089(0.142)	0.063(0.096)	0.073(0.087)	N/A
	SCAD2P.BIC	0.500(0.312)	0.340(0.213)	0.529(0.326)	0.224(0.189)	0.240(0.279)	0.307(0.315)	0.003(0.016)
	SCAD2P.AIC	0.291(0.178)	0.266(0.164)	0.324(0.244)	0.138(0.159)	0.138(0.193)	0.148(0.180)	0.021(0.052)
$K = 7$	ORACLE	0.415(0.247)	0.453(0.233)	0.410(0.250)	0.090(0.144)	0.062(0.091)	0.079(0.100)	N/A
	SCAD2P.BIC	0.509(0.295)	0.350(0.208)	0.465(0.314)	0.218(0.225)	0.207(0.251)	0.284(0.291)	0.001(0.010)
	SCAD2P.AIC	0.316(0.172)	0.291(0.178)	0.324(0.248)	0.159(0.156)	0.159(0.252)	0.170(0.226)	0.017(0.044)
$K = 8$	ORACLE	0.542(0.347)	0.535(0.269)	0.457(0.298)	0.087(0.147)	0.063(0.094)	0.077(0.106)	N/A
	SCAD2P.BIC	0.689(0.408)	0.505(0.491)	0.548(0.356)	0.460(0.520)	0.394(0.338)	0.539(0.460)	0(0)
	SCAD2P.AIC	0.389(0.226)	0.316(0.215)	0.344(0.267)	0.275(0.204)	0.289(0.261)	0.301(0.255)	0.001(0.007)