Supplementary Materials to "Penalized Quantile Regression for Distributed Big Data Using the Slack Variable Representation"

Ye Fan[†], Nan Lin[*] and Xianjun Yin[†]

In this file, we give the closed-form solution of the $\boldsymbol{\beta}$-update in QPADM-slack under the SCAD and MCP penalties and some additional data analysis results.

1 The closed-form solution of $\boldsymbol{\beta}$-update

Note that the SCAD and MCP penalties have the form

$$p_\lambda^{\mathrm{SCAD}}(\beta_j) = \lambda|\beta_j|\mathbb{I}(0 \le |\beta_j| \le \lambda) + \frac{-\beta_j^2 + 2a\lambda|\beta_j| - \lambda^2}{2(a-1)}\mathbb{I}(\lambda < |\beta_j| \le a\lambda) + \frac{(a+1)\lambda^2}{2}\mathbb{I}(|\beta_j| > a\lambda),$$

and

$$p_\lambda^{\mathrm{MCP}}(\beta_j) = \left(\lambda|\beta_j| - \frac{\beta_j^2}{2a}\right)\mathbb{I}(0 \le |\beta_j| \le a\lambda) + \frac{a\lambda^2}{2}\mathbb{I}(|\beta_j| > a\lambda),$$

[*]Nan Lin is corresponding author and Professor, Department of Mathematics and Statistics, Washington University in St. Louis, St. Louis, MO 63130.

[†]Ye Fan is a Ph.D. candidate and Xianjun Yin is Professor, School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 100081, China.

respectively. Denote $\psi_j^k = z_j^k + u_j^k/\rho \ (j = 1, \ldots, p)$, then the solution of the problem (13) in the main article for the SCAD and MCP penalties can be written as follows.

(1) SCAD:

$$\beta_j^{k+1} = \underset{x \in \{x_1, x_2, x_3, 0\}}{\arg\min} \ p_\lambda^{\mathrm{SCAD}}(x) + \frac{\rho}{2}\left(x - \psi_j^k\right)^2,$$

where

$$x_1 = \mathrm{sign}\left(\psi_j^k\right) \min\left(\lambda, \max\left(0, |\psi_j^k| - \lambda/\rho\right)\right),$$

$$x_2 = \mathrm{sign}\left(\psi_j^k\right) \min\left(a\lambda, \max\left(\lambda, \frac{\rho|\psi_j^k|\,(a-1) - a\lambda}{\rho\,(a-1) - 1}\right)\right),$$

and

$$x_3 = \mathrm{sign}\left(\psi_j^k\right) \max\left(a\lambda, |\psi_j^k|\right).$$

(2) MCP:

$$\beta_j^{k+1} = \underset{x \in \{x_1, x_2, 0\}}{\arg\min} \ p_\lambda^{\mathrm{MCP}}(x) + \frac{\rho}{2}\left(x - \psi_j^k\right)^2,$$

where

$$x_1 = \mathrm{sign}\left(\psi_j^k\right) \min\left(a\lambda, \max\left(0, \frac{a\left(\rho|\psi_j^k| - \lambda\right)}{a\rho - 1}\right)\right),$$

and

$$x_2 = \mathrm{sign}\left(\psi_j^k\right) \max\left(a\lambda, |\psi_j^k|\right).$$

2 Additional simulation results

2.1 Simulation results for QPADM-slack with $M = 1$

The main article only showed the performance of QPADM-slack for normal error model under the SCAD penalty. Here we give the simulation results under the MCP penalty and those for Laplace error model. See Tables 1-3.

Table 1: Comparison of QPADM-slack and QPADM for normal error model under the MCP penalty with different data sizes.

(n, p)	Algorithm	τ	Size	P1	P2	AE	Iteration	Time
(20,000, 200)	QPADM-slack	0.3	5(0)	100	100	0.0137(0.0063)	85.81(12.00)	1.28(0.12)
		0.5	4(0)	0	100	0.0058(0.0023)	77.79(10.86)	1.21(0.10)
		0.7	5(0)	100	100	0.0152(0.0073)	83.39(14.06)	1.26(0.13)
	QPADM	0.3	5(0)	100	100	0.0122(0.0052)	219.64(12.62)	2.52(0.12)
		0.5	4(0)	0	100	0.0055(0.0026)	140.64(11.89)	1.77(0.11)
		0.7	5(0)	100	100	0.0128(0.0066)	218.64(12.55)	2.49(0.12)
(50,000, 200)	QPADM-slack	0.3	5(0)	100	100	0.0085(0.0047)	90.85(16.08)	3.35(0.38)
		0.5	4(0)	0	100	0.0034(0.0013)	85.86(13.63)	3.25(0.33)
		0.7	5(0)	100	100	0.0078(0.0038)	92.24(15.52)	3.39(0.37)
	QPADM	0.3	5(0)	100	100	0.0072(0.0030)	233.16(14.70)	6.66(0.35)
		0.5	4(0)	0	100	0.0033(0.0013)	155.99(13.84)	4.85(0.33)
		0.7	5(0)	100	100	0.0075(0.0039)	234.19(12.37)	6.66(0.28)
(100,000, 200)	QPADM-slack	0.3	5(0)	100	100	0.0063(0.0038)	106.28(22.24)	7.53(1.09)
		0.5	4(0)	0	100	0.0024(0.0015)	98.94(16.16)	7.22(0.86)
		0.7	5(0)	100	100	0.0058(0.0034)	111.05(20.87)	7.77(1.09)
	QPADM	0.3	5(0)	100	100	0.0056(0.0029)	253.64(23.90)	14.51(1.66)
		0.5	4(0)	0	100	0.0023(0.0008)	183.57(13.43)	11.15(0.71)
		0.7	5(0)	100	100	0.0058(0.0032)	253.60(24.00)	14.46(1.58)
(50,000, 500)	QPADM-slack	0.3	5(0)	100	100	0.0088(0.0051)	99.04(14.32)	13.23(0.84)
		0.5	4(0)	0	100	0.0039(0.0017)	90.02(14.54)	12.69(0.83)
		0.7	5(0)	100	100	0.0080(0.0035)	99.98(12.01)	13.25(0.69)
	QPADM	0.3	5(0)	100	100	0.0076(0.0033)	242.31(16.99)	21.28(0.98)
		0.5	4(0)	0	100	0.0037(0.0014)	148.57(11.09)	15.93(0.66)
		0.7	5(0)	100	100	0.0077(0.0032)	239.19(14.69)	21.03(0.86)
(50,000, 1000)	QPADM-slack	0.3	5(0)	100	100	0.0095(0.0047)	105.82(17.51)	42.38(2.03)
		0.5	4(0)	0	100	0.0046(0.0017)	95.73(18.57)	41.23(2.12)
		0.7	5(0)	100	100	0.0096(0.0043)	107.14(19.17)	42.51(2.22)
	QPADM	0.3	5(0)	100	100	0.0088(0.0043)	249.25(21.39)	58.58(2.50)
		0.5	4(0)	0	100	0.0047(0.0017)	167.31(20.78)	49.30(2.41)
		0.7	5(0)	100	100	0.0089(0.0043)	249.60(26.32)	58.62(3.02)

[1] The meanings of the notations used in this table are as follows. Size: number of truly selected variables; P1 (%): proportion that x_1 is selected; P2 (%): proportion that x_6, x_{12}, x_{15}, x_{20} are selected; AE: absolute estimation error; Iteration: number of iterations; Time (s): running time. Numbers in the parentheses are the corresponding standard deviations.

Table 2: Comparison of QPADM-slack and QPADM for Laplace error model under the SCAD penalty with different data sizes.

(n, p)	Algorithm	τ	Size	P1	P2	AE	Iteration	Time
(20,000, 200)	QPADM-slack	0.3	8(0)	100	100	0.034(0.011)	118.00(19.39)	1.25(0.12)
		0.5	7(0)	0	100	0.014(0.004)	106.48(16.97)	1.18(0.11)
		0.7	8(0)	100	100	0.033(0.011)	114.16(18.45)	1.26(0.13)
	QPADM	0.3	8(0)	100	100	0.030(0.009)	234.80(25.99)	2.69(0.16)
		0.5	7(0)	0	100	0.013(0.004)	169.60(28.72)	2.07(0.18)
		0.7	8(0)	100	100	0.029(0.009)	222.85(24.04)	2.56(0.15)
(50,000, 200)	QPADM-slack	0.3	8(0)	100	100	0.020(0.007)	128.50(21.02)	4.29(0.51)
		0.5	7(0)	0	100	0.008(0.002)	115.24(18.14)	3.99(0.44)
		0.7	8(0)	100	100	0.020(0.007)	124.96(21.84)	4.21(0.55)
	QPADM	0.3	8(0)	100	100	0.017(0.005)	244.04(24.90)	6.96(0.54)
		0.5	7(0)	0	100	0.007(0.002)	199.03(37.77)	5.90(0.88)
		0.7	8(0)	100	100	0.018(0.006)	232.64(16.11)	6.70(0.39)
(100,000, 200)	QPADM-slack	0.3	8(0)	100	100	0.013(0.005)	140.71(22.80)	9.44(1.18)
		0.5	7(0)	0	100	0.005(0.002)	127.44(20.70)	8.80(1.12)
		0.7	8(0)	100	100	0.014(0.005)	138.49(24.12)	9.28(1.21)
	QPADM	0.3	8(0)	100	100	0.012(0.004)	255.04(26.57)	14.94(1.42)
		0.5	7(0)	0	100	0.005(0.002)	212.56(28.72)	12.85(1.53)
		0.7	8(0)	100	100	0.011(0.005)	253.16(27.43)	14.79(1.46)
(50,000, 500)	QPADM-slack	0.3	8(0)	100	100	0.022(0.007)	134.48(20.77)	15.84(1.36)
		0.5	7(0)	0	100	0.009(0.003)	117.44(15.83)	14.79(1.09)
		0.7	8(0)	100	100	0.021(0.006)	129.89(19.56)	15.55(1.30)
	QPADM	0.3	8(0)	100	100	0.020(0.005)	251.43(26.54)	22.66(1.63)
		0.5	7(0)	0	100	0.008(0.003)	196.11(28.38)	19.40(1.69)
		0.7	8(0)	100	100	0.019(0.006)	244.52(25.28)	22.22(1.66)
(50,000, 1000)	QPADM-slack	0.3	8(0)	100	100	0.024(0.007)	138.98(17.17)	46.16(2.20)
		0.5	7(0)	0	100	0.010(0.003)	118.27(15.88)	43.89(2.73)
		0.7	8(0)	100	100	0.024(0.007)	133.16(20.12)	45.57(2.74)
	QPADM	0.3	8(0)	100	100	0.024(0.007)	261.14(29.20)	60.03(3.37)
		0.5	7(0)	0	100	0.010(0.002)	200.61(23.66)	53.16(2.65)
		0.7	8(0)	100	100	0.023(0.008)	256.38(21.17)	59.52(2.48)

[1] The meanings of the notations used in this table are as follows. Size: number of truly selected variables; P1 (%): proportion that x_1 is selected; P2 (%): proportion that x_2, x_4, x_6, x_9, x_{11}, x_{14} and x_{17} are selected; AE: absolute estimation error; Iteration: number of iterations; Time (s): running time. Numbers in the parentheses are the corresponding standard deviations.

Table 3: Comparison of QPADM-slack and QPADM for Laplace error model under the MCP penalty with different data sizes.

(n, p)	Algorithm	τ	Size	P1	P2	AE	Iteration	Time
(20,000, 200)	QPADM-slack	0.3	8(0)	100	100	0.031(0.010)	111.65(16.48)	1.59(0.17)
		0.5	7(0)	0	100	0.015(0.004)	102.69(18.81)	1.51(0.20)
		0.7	8(0)	100	100	0.033(0.010)	109.92(17.51)	1.58(0.20)
	QPADM	0.3	8(0)	100	100	0.028(0.008)	234.98(26.42)	2.80(0.18)
		0.5	7(0)	0	100	0.014(0.004)	173.09(18.42)	2.17(0.12)
		0.7	8(0)	100	100	0.031(0.009)	221.71(15.06)	2.64(0.10)
(50,000, 200)	QPADM-slack	0.3	8(0)	100	100	0.021(0.006)	127.59(18.30)	4.39(0.50)
		0.5	7(0)	0	100	0.009(0.003)	113.27(17.86)	4.03(0.45)
		0.7	8(0)	100	100	0.019(0.006)	126.76(18.54)	4.35(0.48)
	QPADM	0.3	8(0)	100	100	0.018(0.006)	238.40(22.81)	7.06(0.57)
		0.5	7(0)	0	100	0.008(0.002)	202.62(26.74)	6.18(0.68)
		0.7	8(0)	100	100	0.017(0.006)	229.91(22.12)	6.82(0.54)
(100,000, 200)	QPADM-slack	0.3	8(0)	100	100	0.017(0.005)	137.05(20.18)	9.69(1.26)
		0.5	7(0)	0	100	0.006(0.002)	125.82(20.42)	9.09(1.30)
		0.7	8(0)	100	100	0.017(0.006)	139.05(24.61)	9.74(1.47)
	QPADM	0.3	8(0)	100	100	0.016(0.005)	251.05(25.83)	15.61(1.69)
		0.5	7(0)	0	100	0.005(0.002)	215.36(25.60)	13.70(1.67)
		0.7	8(0)	100	100	0.015(0.005)	249.91(20.08)	15.44(1.31)
(50,000, 500)	QPADM-slack	0.3	8(0)	100	100	0.021(0.007)	135.45(19.80)	15.62(1.34)
		0.5	7(0)	0	100	0.010(0.003)	119.97(18.98)	14.77(1.19)
		0.7	8(0)	100	100	0.022(0.008)	130.35(21.86)	15.31(1.30)
	QPADM	0.3	8(0)	100	100	0.019(0.007)	247.56(26.87)	22.21(1.62)
		0.5	7(0)	0	100	0.009(0.003)	200.46(25.91)	19.43(1.54)
		0.7	8(0)	100	100	0.020(0.007)	239.49(23.41)	21.77(1.51)
(50,000, 1000)	QPADM-slack	0.3	8(0)	100	100	0.023(0.007)	140.69(21.31)	46.46(2.47)
		0.5	7(0)	0	100	0.010(0.003)	122.32(20.07)	44.34(2.28)
		0.7	8(0)	100	100	0.024(0.007)	134.38(20.06)	45.71(2.30)
	QPADM	0.3	8(0)	100	100	0.023(0.007)	266.18(38.37)	60.56(4.34)
		0.5	7(0)	0	100	0.010(0.003)	203.53(22.34)	53.47(2.53)
		0.7	8(0)	100	100	0.023(0.007)	252.72(20.77)	59.07(2.38)

[1] The notations Size, P1, P2, AE, Iteration and Time used here have the same meanings as those in Table 2. Numbers in the parentheses are the corresponding standard deviations.

Besides, to clearly show the difference of QPADM-slack and QPADM in convergence speed, we also plot the convergence path of the objective and the estimation

path of β_6 for a randomly generated data set from the normal error model with $(n, p) = (50,000, 200)$ and $\tau = 0.7$ under the SCAD penalty. See Figure 1.

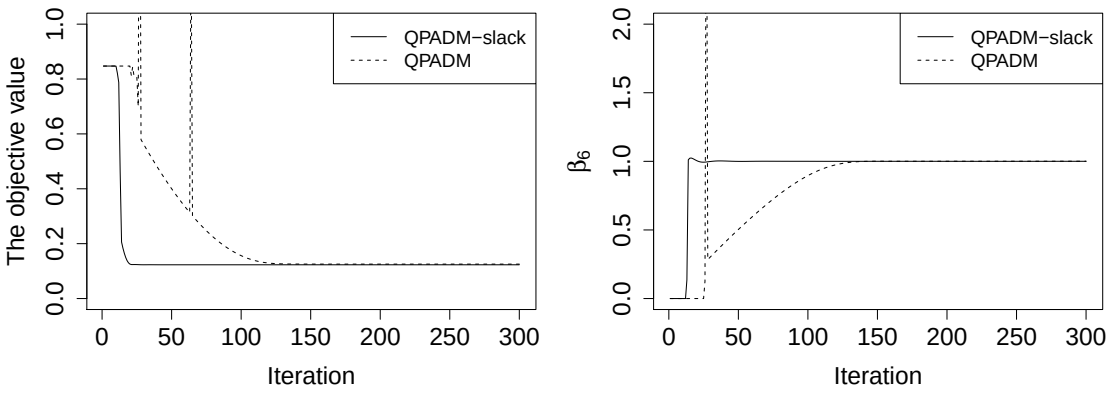

Figure 1: The convergence path of the objective and the estimation path of β_6 for a randomly generated data set with $(n, p) = (50,000, 200)$ and $\tau = 0.7$ under the SCAD penalty.

As mentioned in the main article, the estimation error of QPADM-slack is slightly larger than that of QPADM when the sample size is relatively small ($n = 20,000$). To investigate the causes of this difference, we conduct some additional experiments for the normal error model with different stopping criteria. Remember that the previously used stopping criterion is $\varepsilon = 10^{-3}$. In the new experiments, we reduce it to 5×10^{-4}, 10^{-4}, 5×10^{-5} and 10^{-5}. The simulation results for these cases are presented in Tables 4-5.

Table 4: Comparison of QPADM-slack and QPADM for normal error model under the SCAD penalty with different stopping criteria.

τ	QPADM-slack			QPADM		
	AE	Iteration	Time	AE	Iteration	Time
			$\varepsilon = 5 \times 10^{-4}$			
0.3	0.0139(0.0065)	87.64(15.34)	1.32(0.15)	0.0121(0.0042)	250.69(33.24)	2.83(0.31)
0.5	0.0056(0.0022)	82.50(14.84)	1.27(0.15)	0.0053(0.0022)	168.85(27.64)	2.07(0.26)
0.7	0.0140(0.0067)	91.17(18.59)	1.35(0.18)	0.0127(0.0056)	244.24(23.91)	2.76(0.23)
			$\varepsilon = 10^{-4}$			
0.3	0.0132(0.0050)	124.96(30.61)	1.76(0.32)	0.0120(0.0052)	345.65(97.12)	3.93(1.01)
0.5	0.0055(0.0022)	117.00(35.66)	1.68(0.37)	0.0050(0.0020)	262.47(106.44)	3.07(1.01)
0.7	0.0138(0.0065)	128.53(37.30)	1.81(0.39)	0.0124(0.0065)	334.99(93.78)	3.80(0.95)
			$\varepsilon = 5 \times 10^{-5}$			
0.3	0.0134(0.0053)	145.26(40.34)	1.84(0.38)	0.0118(0.0060)	392.06(104.09)	4.08(0.96)
0.5	0.0053(0.0021)	145.44(49.32)	1.86(0.47)	0.0049(0.0019)	339.07(125.00)	3.60(1.15)
0.7	0.0138(0.0064)	148.10(48.11)	1.87(0.45)	0.0123(0.0065)	382.32(105.57)	4.00(0.98)
			$\varepsilon = 10^{-5}$			
0.3	0.0133(0.0051)	254.75(112.43)	2.87(1.04)	0.0116(0.0051)	474.11(67.41)	4.85(0.63)
0.5	0.0053(0.0021)	252.55(93.96)	2.86(0.88)	0.0048(0.0018)	467.52(74.64)	4.79(0.68)
0.7	0.0137(0.0065)	251.01(92.92)	2.83(0.87)	0.0120(0.0066)	459.45(83.38)	4.72(0.77)

[1] In this experiment, we stop the algorithm if the change of the objective value at two successive iterations is less than ε or the number of iterations exceeds 1000. The notations AE, Iteration and Time used here have the same meanings as those in Table 1. Numbers in the parentheses are the corresponding standard deviations.

Table 5: Comparison of QPADM-slack and QPADM for normal error model under the MCP penalty with different stopping criteria.

τ	QPADM-slack			QPADM		
	AE	Iteration	Time	AE	Iteration	Time
			$\varepsilon = 5 \times 10^{-4}$			
0.3	0.0131(0.0056)	88.61(14.16)	1.30(0.13)	0.0121(0.0052)	250.94(34.23)	2.78(0.31)
0.5	0.0056(0.0022)	83.20(16.06)	1.26(0.15)	0.0054(0.0024)	171.16(39.47)	2.05(0.36)
0.7	0.0142(0.0068)	90.34(18.31)	1.32(0.17)	0.0127(0.0066)	244.48(24.22)	2.71(0.22)
			$\varepsilon = 10^{-4}$			
0.3	0.0131(0.0049)	122.13(34.38)	1.63(0.32)	0.0119(0.0052)	353.55(134.19)	3.75(1.24)
0.5	0.0056(0.0022)	119.23(33.92)	1.60(0.32)	0.0049(0.0021)	282.55(168.72)	3.09(1.55)
0.7	0.0137(0.0068)	124.46(37.60)	1.65(0.36)	0.0123(0.0065)	373.25(157.86)	3.93(1.45)
			$\varepsilon = 5 \times 10^{-5}$			
0.3	0.0132(0.0051)	143.87(49.29)	1.83(0.45)	0.0118(0.0050)	460.23(223.35)	4.75(2.11)
0.5	0.0055(0.0021)	149.24(53.83)	1.88(0.51)	0.0048(0.0019)	403.60(247.43)	4.20(2.27)
0.7	0.0138(0.0067)	156.60(48.44)	1.95(0.46)	0.0121(0.0066)	481.42(243.12)	4.93(2.26)
			$\varepsilon = 10^{-5}$			
0.3	0.0133(0.0051)	257.50(112.86)	2.97(1.09)	0.0112(0.0039)	808.25(258.28)	8.11(2.43)
0.5	0.0055(0.0021)	245.61(99.59)	2.86(0.95)	0.0043(0.0015)	758.62(295.22)	7.71(2.78)
0.7	0.0137(0.0064)	265.27(103.11)	3.03(0.99)	0.0114(0.0057)	809.42(258.63)	8.11(2.42)

[1] In this experiment, we stop the algorithm if the change of the objective value at two successive iterations is less than ε or the number of iterations exceeds 1000. The notations AE, Iteration and Time used here have the same meanings as those in Table 1. Numbers in the parentheses are the corresponding standard deviations.

Besides, the simulation results presented in the main article also show that, QPADM-slack and QPADM perform similarly in model selection accuracy for the previously considered model, and can always successfully select the true model. To investigate whether the two algorithms have any difference in model selection accuracy for more challenging situations, we reduce the true coefficient of the covariate x_6, and set β_6 to be 0.5, 0.1 or 0.05. The simulation results for these cases are shown in Tables 6-7.

Table 6: Comparison of QPADM-slack and QPADM for normal error model under the SCAD penalty with different coefficients for the covariate x_6.

Algorithm	τ	Size	P1	P2	AE	Iteration	Time
				$\beta_6 = 0.5$			
QPADM-slack	0.3	5.00(0.00)	100	100	0.0089(0.0047)	89.73(15.74)	3.40(0.44)
	0.5	4.00(0.00)	0	100	0.0034(0.0014)	83.62(14.83)	3.25(0.41)
	0.7	5.00(0.00)	100	100	0.0078(0.0047)	89.29(18.99)	3.39(0.48)
QPADM	0.3	5.00(0.00)	100	100	0.0083(0.0039)	219.96(14.79)	6.46(0.43)
	0.5	4.00(0.00)	0	100	0.0033(0.0012)	145.63(12.16)	4.71(0.40)
	0.7	5.00(0.00)	100	100	0.0075(0.0039)	220.72(12.82)	6.48(0.47)
				$\beta_6 = 0.1$			
QPADM-slack	0.3	5.00(0.00)	100	100	0.0084(0.0036)	89.27(16.39)	3.35(0.40)
	0.5	4.00(0.00)	0	100	0.0033(0.0014)	79.59(13.70)	3.16(0.36)
	0.7	5.00(0.00)	100	100	0.0077(0.0040)	87.75(15.93)	3.35(0.41)
QPADM	0.3	5.00(0.00)	100	100	0.0083(0.0040)	214.18(13.64)	6.31(0.41)
	0.5	4.00(0.00)	0	100	0.0033(0.0012)	137.14(10.68)	4.48(0.29)
	0.7	5.00(0.00)	100	100	0.0075(0.0039)	214.00(12.99)	6.33(0.45)
				$\beta_6 = 0.05$			
QPADM-slack	0.3	4.96(0.20)	100	96	0.0106(0.0140)	91.39(16.72)	3.62(0.58)
	0.5	3.96(0.20)	0	96	0.0053(0.0108)	81.71(13.71)	3.35(0.42)
	0.7	4.97(0.17)	100	97	0.0094(0.0118)	91.28(14.49)	3.59(0.41)
QPADM	0.3	4.98(0.14)	100	98	0.0093(0.0090)	212.73(13.71)	6.59(0.53)
	0.5	3.94(0.24)	0	94	0.0062(0.0120)	136.22(10.24)	4.70(0.38)
	0.7	4.96(0.20)	100	96	0.0097(0.0103)	215.51(15.13)	6.62(0.49)

[1] In this experiment, we set a smaller true value for β_6. The notations Size, P1, P2, AE, Iteration and Time used here also have the same meanings as those in Table 1. Numbers in the parentheses are the corresponding standard deviations.

Table 7: Comparison of QPADM-slack and QPADM for normal error model under the MCP penalty with different coefficients for the covariate x_6.

Algorithm	τ	Size	P1	P2	AE	Iteration	Time
				β_6 = 0.5			
QPADM-slack	0.3	5.00(0.00)	100	100	0.0086(0.0040)	88.76(15.41)	3.27(0.36)
	0.5	4.00(0.00)	0	100	0.0034(0.0015)	83.95(12.70)	3.16(0.30)
	0.7	5.00(0.00)	100	100	0.0076(0.0040)	88.50(18.00)	3.26(0.42)
QPADM	0.3	5.00(0.00)	100	100	0.0083(0.0039)	219.96(14.79)	6.26(0.34)
	0.5	4.00(0.00)	0	100	0.0033(0.0012)	145.02(12.56)	4.53(0.29)
	0.7	5.00(0.00)	100	100	0.0075(0.0039)	220.72(12.82)	6.27(0.30)
				β_6 = 0.1			
QPADM-slack	0.3	5.00(0.00)	100	100	0.0083(0.0038)	89.04(16.43)	3.48(0.56)
	0.5	4.00(0.00)	0	100	0.0035(0.0018)	79.18(14.86)	3.22(0.45)
	0.7	5.00(0.00)	100	100	0.0072(0.0033)	89.28(14.76)	3.44(0.41)
QPADM	0.3	5.00(0.00)	100	100	0.0083(0.0039)	215.72(15.02)	6.51(0.61)
	0.5	4.00(0.00)	0	100	0.0032(0.0012)	135.83(12.08)	4.57(0.39)
	0.7	5.00(0.00)	100	100	0.0075(0.0039)	216.80(13.51)	6.47(0.46)
				β_6 = 0.05			
QPADM-slack	0.3	4.97(0.17)	100	97	0.0095(0.0102)	90.87(15.72)	3.39(0.44)
	0.5	3.97(0.17)	0	97	0.0051(0.0097)	80.51(14.42)	3.13(0.38)
	0.7	4.98(0.14)	100	98	0.0084(0.0091)	92.80(14.43)	3.45(0.42)
QPADM	0.3	4.98(0.14)	100	98	0.0094(0.0099)	212.24(13.48)	6.22(0.45)
	0.5	3.93(0.26)	0	93	0.0067(0.0128)	136.01(10.56)	4.41(0.34)
	0.7	4.96(0.20)	100	96	0.0096(0.0102)	217.53(15.50)	6.31(0.49)

[1] In this experiment, we set a smaller true value for β_6. The notations Size, P1, P2, AE, Iteration and Time used here also have the same meanings as those in Table 1. Numbers in the parentheses are the corresponding standard deviations.

2.2 Simulation results for parallel QPADM-slack

Due to space limit, in the main article we did not show the specific results for parallel QPADM-slack, but only presented two plots for its computational time and estimation error versus M under normal and Laplace error models. Here we give the results for the two models under the SCAD and MCP penalties in Tables 8-11.

Table 8: Comparison of QPADM-slack and QPADM for normal error model under the SCAD penalty with different number of partitions.

Algorithm	M	Size	P1	P2	AE	Iteration	Time
				$\tau = 0.3$			
QPADM-slack	1	5.00(0.00)	100	100	0.0035(0.0017)	185.49(28.54)	25.71(3.74)
	10	5.00(0.00)	100	100	0.0038(0.0018)	209.96(29.97)	2.78(0.36)
	20	5.00(0.00)	100	100	0.0041(0.0019)	217.09(32.11)	1.53(0.32)
	50	5.06(0.24)	100	100	0.0055(0.0023)	263.76(52.67)	0.90(0.24)
	100	5.40(0.57)	100	100	0.0065(0.0033)	328.61(96.75)	0.67(0.21)
QPADM	1	5.00(0.00)	100	100	0.0034(0.0016)	318.48(22.76)	40.70(2.97)
	10	5.00(0.00)	100	100	0.0037(0.0017)	338.39(25.42)	4.06(0.30)
	20	5.00(0.00)	100	100	0.0039(0.0019)	344.22(23.84)	2.14(0.21)
	50	5.07(0.29)	100	100	0.0052(0.0019)	365.72(43.46)	1.08(0.19)
	100	5.34(0.54)	100	100	0.0063(0.0025)	426.93(78.80)	0.75(0.18)
				$\tau = 0.5$			
QPADM-slack	1	4.00(0.00)	0	100	0.0015(0.0007)	160.85(22.90)	23.03(2.96)
	10	4.00(0.00)	0	100	0.0016(0.0008)	181.70(26.95)	2.49(0.31)
	20	4.00(0.00)	0	100	0.0019(0.0008)	190.14(29.85)	1.37(0.30)
	50	4.01(0.10)	0	100	0.0024(0.0009)	224.82(38.88)	0.72(0.17)
	100	4.10(0.30)	0	100	0.0028(0.0018)	270.04(71.46)	0.48(0.17)
QPADM	1	4.00(0.00)	0	100	0.0015(0.0006)	213.07(19.57)	28.67(2.56)
	10	4.00(0.00)	0	100	0.0016(0.0007)	232.51(22.05)	2.94(0.26)
	20	4.00(0.00)	0	100	0.0017(0.0007)	238.76(18.07)	1.58(0.18)
	50	4.01(0.10)	0	100	0.0027(0.0012)	261.13(38.32)	0.80(0.17)
	100	4.14(0.40)	0	100	0.0030(0.0026)	313.17(59.93)	0.56(0.15)
				$\tau = 0.7$			
QPADM-slack	1	5.00(0.00)	100	100	0.0039(0.0017)	177.78(23.89)	24.67(3.07)
	10	5.00(0.00)	100	100	0.0040(0.0019)	200.40(21.93)	2.68(0.29)
	20	5.00(0.00)	100	100	0.0040(0.0021)	214.11(26.01)	1.51(0.24)
	50	5.05(0.22)	100	100	0.0056(0.0022)	271.77(61.46)	0.93(0.26)
	100	5.31(0.46)	100	100	0.0061(0.0032)	339.35(92.88)	0.68(0.21)
QPADM	1	5.00(0.00)	100	100	0.0038(0.0017)	317.01(22.48)	40.20(2.92)
	10	5.00(0.00)	100	100	0.0039(0.0018)	339.47(26.42)	4.08(0.35)
	20	5.00(0.00)	100	100	0.0040(0.0019)	343.40(25.22)	2.16(0.24)
	50	5.03(0.17)	100	100	0.0054(0.0023)	365.02(40.58)	1.06(0.18)
	100	5.41(0.57)	100	100	0.0064(0.0026)	423.75(71.86)	0.74(0.17)

[1] The meanings of the notations used in this table are as follows. M: number of partitions. Size: number of truly selected variables; P1 (%): proportion that x_1 is selected; P2 (%): proportion that x_6, x_{12}, x_{15} and x_{20} are selected; AE: absolute estimation error; Iteration: number of iterations; Time (s): running time. Numbers in the parentheses are the corresponding standard deviations.

Table 9: Comparison of QPADM-slack and QPADM for normal error model under the MCP penalty with different number of partitions.

Algorithm	M	Size	P1	P2	AE	Iteration	Time
				$\tau = 0.3$			
QPADM-slack	1	5.00(0.00)	100	100	0.0035(0.0016)	190.24(27.12)	25.67(3.49)
	10	5.00(0.00)	100	100	0.0037(0.0017)	216.58(33.40)	2.78(0.39)
	20	5.00(0.00)	100	100	0.0039(0.0015)	223.42(40.74)	1.50(0.29)
	50	5.05(0.22)	100	100	0.0055(0.0021)	277.44(77.09)	0.87(0.25)
	100	5.31(0.46)	100	100	0.0063(0.0029)	361.38(82.55)	0.69(0.22)
QPADM	1	5.00(0.00)	100	100	0.0034(0.0015)	318.98(20.99)	39.71(2.58)
	10	5.00(0.00)	100	100	0.0035(0.0016)	341.73(22.48)	4.01(0.26)
	20	5.00(0.00)	100	100	0.0038(0.0017)	348.12(25.87)	2.10(0.18)
	50	5.06(0.24)	100	100	0.0055(0.0017)	366.52(48.87)	1.03(0.16)
	100	5.41(0.57)	100	100	0.0065(0.0023)	425.30(61.10)	0.71(0.16)
				$\tau = 0.5$			
QPADM-slack	1	4.00(0.00)	0	100	0.0015(0.0009)	160.44(23.50)	22.93(3.10)
	10	4.00(0.00)	0	100	0.0018(0.0010)	183.87(28.62)	2.45(0.31)
	20	4.00(0.00)	0	100	0.0019(0.0010)	192.32(30.32)	1.33(0.24)
	50	4.01(0.10)	0	100	0.0024(0.0010)	233.31(60.26)	0.75(0.23)
	100	4.08(0.27)	0	100	0.0025(0.0016)	274.27(68.79)	0.53(0.17)
QPADM	1	4.00(0.00)	0	100	0.0015(0.0006)	214.74(16.38)	28.47(2.12)
	10	4.00(0.00)	0	100	0.0017(0.0007)	230.87(24.56)	2.88(0.26)
	20	4.00(0.00)	0	100	0.0018(0.0007)	239.82(25.53)	1.54(0.20)
	50	4.02(0.14)	0	100	0.0029(0.0011)	266.76(42.96)	0.79(0.17)
	100	4.11(0.31)	0	100	0.0029(0.0025)	319.13(59.71)	0.55(0.15)
				$\tau = 0.7$			
QPADM-slack	1	5.00(0.00)	100	100	0.0036(0.0019)	186.60(28.96)	25.26(3.74)
	10	5.00(0.00)	100	100	0.0039(0.0017)	210.56(30.27)	2.71(0.33)
	20	5.00(0.00)	100	100	0.0042(0.0016)	223.41(35.06)	1.51(0.31)
	50	5.03(0.17)	100	100	0.0053(0.0023)	270.09(73.88)	0.85(0.28)
	100	5.28(0.45)	100	100	0.0058(0.0028)	352.63(80.86)	0.65(0.20)
QPADM	1	5.00(0.00)	100	100	0.0035(0.0016)	317.65(21.61)	39.57(2.81)
	10	5.00(0.00)	100	100	0.0037(0.0016)	340.88(23.40)	4.02(0.26)
	20	5.00(0.00)	100	100	0.0038(0.0016)	343.65(25.19)	2.09(0.22)
	50	5.03(0.17)	100	100	0.0052(0.0021)	359.03(51.06)	1.03(0.19)
	100	5.30(0.46)	100	100	0.0061(0.0023)	434.48(60.05)	0.72(0.15)

[1] The notations M, Size, P1, P2, AE, Iteration and Time used here have the same meanings as those in Table 8. Numbers in the parentheses are the corresponding standard deviations.

Table 10: Comparison of QPADM-slack and QPADM for Laplace error model under the SCAD penalty with different number of partitions.

Algorithm	M	Size	P1	P2	AE	Iteration	Time
				$\tau = 0.3$			
QPADM-slack	1	8.00(0.00)	100	100	0.0076(0.0024)	253.06(35.08)	31.46(4.08)
	10	8.00(0.00)	100	100	0.0085(0.0026)	337.13(65.85)	3.96(0.70)
	20	8.00(0.00)	100	100	0.0092(0.0032)	356.65(70.27)	2.16(0.46)
	50	8.01(0.10)	100	100	0.0105(0.0030)	371.74(68.65)	1.01(0.18)
	100	8.13(0.42)	100	100	0.0124(0.0028)	424.59(82.11)	0.69(0.21)
QPADM	1	8.00(0.00)	100	100	0.0075(0.0025)	337.91(35.14)	40.82(4.21)
	10	8.00(0.00)	100	100	0.0082(0.0027)	402.62(58.54)	4.56(0.64)
	20	8.00(0.00)	100	100	0.0095(0.0031)	447.48(52.20)	2.54(0.34)
	50	8.05(0.22)	100	100	0.0104(0.0027)	460.24(47.45)	1.17(0.14)
	100	8.09(0.29)	100	100	0.0123(0.0033)	508.34(75.39)	0.74(0.19)
				$\tau = 0.5$			
QPADM-slack	1	7.00(0.00)	0	100	0.0032(0.0009)	212.49(28.25)	26.98(3.52)
	10	7.00(0.00)	0	100	0.0042(0.0012)	266.86(40.83)	3.22(0.43)
	20	7.00(0.00)	0	100	0.0046(0.0013)	282.76(51.65)	1.75(0.33)
	50	7.01(0.10)	0	100	0.0048(0.0016)	314.38(66.82)	0.87(0.17)
	100	7.02(0.14)	0	100	0.0050(0.0024)	366.32(79.63)	0.61(0.20)
QPADM	1	7.00(0.00)	0	100	0.0031(0.0009)	261.70(28.20)	32.49(3.46)
	10	7.00(0.00)	0	100	0.0039(0.0012)	316.42(30.46)	3.66(0.30)
	20	7.00(0.00)	0	100	0.0043(0.0016)	340.21(31.34)	1.98(0.22)
	50	7.04(0.20)	1	100	0.0051(0.0018)	367.07(42.53)	0.94(0.13)
	100	7.05(0.22)	2	100	0.0052(0.0035)	412.27(46.33)	0.66(0.12)
				$\tau = 0.7$			
QPADM-slack	1	8.00(0.00)	100	100	0.0080(0.0031)	243.82(32.70)	30.27(4.01)
	10	8.00(0.00)	100	100	0.0099(0.0034)	304.82(55.41)	3.62(0.57)
	20	8.00(0.00)	100	100	0.0103(0.0033)	335.22(59.68)	2.00(0.38)
	50	8.01(0.10)	100	100	0.0110(0.0035)	350.91(59.53)	0.94(0.16)
	100	8.08(0.31)	100	100	0.0117(0.0045)	406.12(77.00)	0.66(0.20)
QPADM	1	8.00(0.00)	100	100	0.0078(0.0032)	313.76(29.60)	38.19(3.92)
	10	8.00(0.00)	100	100	0.0101(0.0037)	380.20(45.72)	4.28(0.43)
	20	8.00(0.00)	100	100	0.0103(0.0039)	421.69(46.00)	2.39(0.30)
	50	8.02(0.14)	100	100	0.0106(0.0032)	457.26(49.52)	1.15(0.14)
	100	8.14(0.49)	100	100	0.0123(0.0036)	498.53(67.71)	0.73(0.17)

[1] The meanings of the notations used in this table are as follows. M: number of partitions. Size: number of truly selected variables; P1 (%): proportion that x_1 is selected; P2 (%): proportion that x_2, x_4, x_6, x_9, x_{11}, x_{14} and x_{17} are selected; AE: absolute estimation error; Iteration: number of iterations; Time (s): running time. Numbers in the parentheses are the corresponding standard deviations.

Table 11: Comparison of QPADM-slack and QPADM for Laplace error model under the MCP penalty with different number of partitions.

Algorithm	M	Size	P1	P2	AE	Iteration	Time
				$\tau = 0.3$			
QPADM-slack	1	8.00(0.00)	100	100	0.0076(0.0030)	272.66(41.18)	33.99(4.76)
	10	8.00(0.00)	100	100	0.0084(0.0031)	332.17(59.98)	3.94(0.67)
	20	8.00(0.00)	100	100	0.0091(0.0024)	350.85(73.42)	2.13(0.46)
	50	8.01(0.10)	100	100	0.0105(0.0033)	360.35(78.41)	0.98(0.21)
	100	8.09(0.29)	100	100	0.0116(0.0036)	407.69(81.11)	0.67(0.21)
QPADM	1	8.00(0.00)	100	100	0.0075(0.0029)	334.98(28.76)	40.51(3.51)
	10	8.00(0.00)	100	100	0.0090(0.0029)	403.10(60.68)	4.57(0.70)
	20	8.00(0.00)	100	100	0.0100(0.0022)	449.53(55.80)	2.55(0.35)
	50	8.03(0.22)	100	100	0.0106(0.0034)	462.18(56.18)	1.18(0.18)
	100	8.11(0.34)	100	100	0.0126(0.0038)	498.29(67.22)	0.73(0.17)
				$\tau = 0.5$			
QPADM-slack	1	7.00(0.00)	0	100	0.0032(0.0010)	223.68(28.79)	28.45(3.44)
	10	7.00(0.00)	0	100	0.0042(0.0013)	263.20(43.29)	3.18(0.45)
	20	7.00(0.00)	0	100	0.0045(0.0013)	296.38(50.51)	1.84(0.33)
	50	7.00(0.00)	0	100	0.0045(0.0016)	307.26(54.52)	0.85(0.15)
	100	7.06(0.24)	3	100	0.0056(0.0024)	377.72(71.89)	0.63(0.18)
QPADM	1	7.00(0.00)	0	100	0.0032(0.0010)	262.40(25.46)	32.56(3.42)
	10	7.00(0.00)	0	100	0.0042(0.0018)	314.08(34.96)	3.64(0.33)
	20	7.00(0.00)	0	100	0.0045(0.0016)	344.74(39.76)	2.00(0.24)
	50	7.00(0.00)	0	100	0.0045(0.0016)	366.11(41.71)	0.93(0.13)
	100	7.08(0.27)	4	100	0.0058(0.0028)	421.98(41.13)	0.67(0.11)
				$\tau = 0.7$			
QPADM-slack	1	8.00(0.00)	100	100	0.0078(0.0027)	249.83(34.93)	31.16(4.03)
	10	8.00(0.00)	100	100	0.0095(0.0026)	313.89(62.30)	3.74(0.71)
	20	8.00(0.00)	100	100	0.0101(0.0029)	333.24(66.92)	2.03(0.44)
	50	8.01(0.10)	100	100	0.0104(0.0031)	361.04(80.61)	0.99(0.22)
	100	8.08(0.27)	100	100	0.0115(0.0036)	405.97(75.00)	0.67(0.19)
QPADM	1	8.00(0.00)	100	100	0.0075(0.0026)	341.00(30.08)	41.39(3.53)
	10	8.00(0.00)	100	100	0.0097(0.0030)	403.29(59.18)	4.63(0.67)
	20	8.00(0.00)	100	100	0.0097(0.0030)	421.94(42.74)	2.46(0.28)
	50	8.01(0.10)	100	100	0.0102(0.0032)	456.92(52.84)	1.18(0.15)
	100	8.09(0.32)	100	100	0.0114(0.0035)	502.91(64.42)	0.72(0.15)

[1] The notations M, Size, P1, P2, AE, Iteration and Time used here have the same meanings as those in Table 10. Numbers in the parentheses are the corresponding standard deviations.

2.3 Simulation results for QPADM-slack with $n < p$

In Section 4 of the main article, we mainly focused on the comparison of QPADM-slack and QPADM in the cases $n > p$. Actually, we also did some attempts for investigating the performance of QPADM-slack in short-and-fat data. Although QPADM-slack still converges faster than QPADM in this case, it may bring inflated estimation error. Here we present the simulation results in the case $(n, p) = (500, 1000)$ for normal error model in Table 12.

Table 12: Comparison of QPADM-slack and QPADM for normal error model under the SCAD and MCP penalties with $(n, p) = (500, 1000)$.

Algorithm	τ	Size	P1	P2	AE	Iteration	Time(s)
				SCAD			
QPADM-slack	0.3	6.61(2.17)	99	100	0.18(0.13)	267.62(13.08)	1.78(0.06)
	0.5	4.26(0.60)	0	100	0.16(0.10)	251.74(11.28)	1.76(0.05)
	0.7	6.36(1.80)	98	100	0.21(0.18)	247.20(11.65)	1.76(0.06)
QPADM	0.3	6.27(1.04)	100	100	0.15(0.07)	985.73(47.04)	3.27(0.28)
	0.5	4.83(0.90)	0	100	0.07(0.04)	993.77(40.94)	3.29(0.31)
	0.7	5.73(0.90)	100	100	0.14(0.05)	1102.29(43.62)	3.49(0.26)
				MCP			
QPADM-slack	0.3	6.50(1.55)	100	100	0.20(0.15)	252.98(11.09)	1.76(0.05)
	0.5	4.25(0.50)	0	100	0.15(0.11)	254.71(10.40)	1.77(0.04)
	0.7	6.02(1.40)	99	100	0.21(0.19)	267.43(10.55)	1.79(0.05)
QPADM	0.3	6.39(1.52)	100	100	0.15(0.06)	971.92(43.12)	3.25(0.22)
	0.5	4.84(0.58)	0	100	0.08(0.04)	944.87(35.88)	3.17(0.17)
	0.7	5.55(0.77)	100	100	0.15(0.05)	1074.50(47.31)	3.41(0.16)

[1] The notations Size, P1, P2, AE, Iteration and Time used here have the same meanings as those in Table 1. Numbers in the parentheses are the corresponding standard deviations.

3 Real data analysis under the MCP penalty

In the main article, we reported the results of the real data analysis under the SCAD penalty. Actually, we also modeled the news popularity data with the MCP penalized quantile regression. The results of this case are presented in Table 13.

Table 13: Analysis of the news popularity data under the MCP penalty.

M	Variable		Nonzero	PE	Iteration	Time(s)
	x_{14}	x_{27}				
QPADM-slack ($\tau = 0.5$)						
5	-0.020(0.001)	0.052(0.001)	34.85(1.45)	0.10(0.01)	116.69(9.80)	0.17(0.03)
20	-0.019(0.001)	0.054(0.001)	35.04(1.37)	0.10(0.01)	135.21(19.16)	0.05(0.02)
100	-0.018(0.001)	0.054(0.001)	37.52(1.52)	0.10(0.01)	179.75(28.09)	0.02(0.01)
QPADM ($\tau = 0.5$)						
5	-0.022(0.002)	0.055(0.001)	33.88(2.13)	0.10(0.01)	212.39(43.98)	0.28(0.06)
20	-0.022(0.002)	0.056(0.001)	35.29(1.78)	0.10(0.01)	214.23(38.71)	0.08(0.02)
100	-0.023(0.002)	0.058(0.001)	36.04(2.25)	0.10(0.01)	272.69(39.44)	0.04(0.01)
QPADM-slack ($\tau = 0.9$)						
5	-0.126(0.009)	0.289(0.007)	38.63(1.44)	0.12(0.01)	460.47(52.56)	0.60(0.11)
20	-0.124(0.008)	0.291(0.006)	38.51(1.45)	0.12(0.01)	452.23(34.69)	0.16(0.04)
100	-0.125(0.007)	0.295(0.006)	38.50(1.23)	0.12(0.01)	465.42(25.44)	0.05(0.03)
QPADM ($\tau = 0.9$)						
5	-0.139(0.012)	0.316(0.008)	39.67(2.06)	0.12(0.01)	592.77(64.57)	0.77(0.18)
20	-0.136(0.018)	0.314(0.008)	41.13(1.77)	0.12(0.01)	627.45(112.49)	0.21(0.06)
100	-0.125(0.016)	0.317(0.008)	41.40(1.46)	0.12(0.01)	609.74(102.56)	0.07(0.04)

[1] The meanings of the notations used in this table are as follows. Variable: estimation coefficient for the reported variable; Nonzero: number of nonzero coefficients selected; AE: prediction error; Iteration: number of iterations; Time: running time. Numbers in the parentheses are the corresponding standard deviations.