Supplementary Material to "Bayesian Model Selection for a Linear Model with Grouped Covariates"

Xiaoyi Min and Dongchu Sun [1]

In this supplementary document, we present another simulation study which examines the main effect in a 2-way ANOVA model. The models being considered are M_3 and M_1 in Example 1. For 'TESS', the same priors as in Section 5 are applied to different components of $\boldsymbol{g}$.

Data Generation: Without loss of generality, μ is set as 0, and σ^2 is set as 1. p_1 and p_2 are set to 3, 5, 10, 20, or 50, and k is set to 3, 5, or 20. $\boldsymbol{\beta}_1$ is generated from $N_{p_1}(\mathbf{0}, \tilde{g}_1\sigma^2\boldsymbol{I}_{p_1})$, $\boldsymbol{\beta}_2$ is generated from $N_{p_2}(\mathbf{0}, \tilde{g}_2\sigma^2\boldsymbol{I}_{p_2})$, and $\boldsymbol{\epsilon}$ is generated from $N_n(\mathbf{0}, \sigma^2\boldsymbol{I}_{p_1p_2k})$, where $\tilde{g}_1$ is fixed at 1, and $\tilde{g}_2$ assumes values 0, 0.1, and 0.3. Under each setting, we calculate the data $\boldsymbol{y}$ according to M_3 and repeat for 100 times . $\ln(B_{31})$ is calculated for the 100 groups of data under the three priors, and the means and the standard deviations of $\ln(B_{31})$ are summarized under each prior in Tables 1–9. Interpretation: When M_1 is the true model (*i.e.* $\tilde{g}_2 = 0$), all three methods suggest the correct model with negative averages of $\ln(B_{31})$. The absolute values of $\ln(B_{31})$ from the three methods are close for small sample sizes, and they increase as the sample size increases.

When $\tilde{g}_2 \neq 0$, 'TESS' gives better results than the other two methods as it yields greater values of $\ln(B_{31})$ and it always leads to the correct choice of model except for the smallest sample size. Also note that for small $\tilde{g}_2$, if p_1 and k are fixed at small values, $\ln(B_{31})$ given by methods 'BG' and 'IZS' sometimes decreases as p_2 increases, which suggests that the Bayes factor is inconsistent. Since 'BG' and 'IZS' do not always lead to consistent Bayes factor, our recommended prior for comparing models M_3 and M_1 is again 'TESS'. However, it should be used with cautious since it could select the wrong model when p_1, p_2, and k are small and the effect of $\boldsymbol{\beta}_2$ is weak.

[1]Xiaoyi Min is Postdoctoral Associate (Email: xiaoyi.min@yale.edu), Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut 06511; and Dongchu Sun is Professor (Email: sund@missouri.edu), Department of Statistics, University of Missouri, Columbia, Missouri 65211. This research was supported in part by NSF Grants DMS-1007874 SES-1024080, SES-1260806, and NIH Grant R01DA016750. Most of the work was done while Xiaoyi Min was a graduate student at the University of Missouri.

Table 1: Mean and Standard Deviation of $\ln(B_{31})$: $k = 3$, $\tilde{g}_2$=0

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	-1.76(1.34)	-1.97(1.08)	-2.06(1.11)	-2.09(0.92)	-2.86(2.59)
	IZS	-2.32(1.3)	-2.9(1.1)	-3.55(1.12)	-4.34(1.1)	-8.66(3.77)
	TESS	-1.51(1.14)	-2.01(0.96)	-2.62(1.01)	-3.24(0.84)	-4.04(0.94)
$p_2 = 5$	BG	-4.03(1.51)	-4.29(1.34)	-4.6(1.67)	-4.77(1.34)	-6.23(2.88)
	IZS	-4.65(1.46)	-5.6(1.29)	-6.85(1.69)	-8.35(1.51)	-13.81(4.21)
	TESS	-2.4(1.17)	-3.1(1.07)	-4.12(1.47)	-5.27(1.28)	-6.97(1.56)
$p_2 = 10$	BG	-10.25(2.99)	-11.36(2.6)	-12.33(2.41)	-13.33(2.28)	-16.63(3.64)
	IZS	-10.31(2.31)	-12.55(2.34)	-15.51(2.33)	-18.61(2.3)	-26.51(4.87)
	TESS	-3.68(1.63)	-4.98(1.78)	-6.95(1.98)	-9.22(1.9)	-13.05(2.05)
$p_2 = 20$	BG	-23.08(4.13)	-26.17(3.97)	-30.41(3.27)	-33.25(3.54)	-40.63(4.78)
	IZS	-22.93(3.58)	-27.57(3.38)	-33.89(3.16)	-40.53(3.55)	-52.1(5.46)
	TESS	-5.46(1.93)	-7.42(2.21)	-10.74(2.13)	-14.97(2.51)	-22.52(3.39)
$p_2 = 50$	BG	-67.84(6.59)	-80.11(6.4)	-93.2(6.42)	-106.12(6.17)	-126.15(8.32)
	IZS	-71.22(7.11)	-84.39(6.79)	-100.17(6.6)	-116.83(6.72)	-143.96(7.98)
	TESS	-8.85(2.6)	-12.87(2.78)	-19.81(4.15)	-30.76(4.48)	-47.64(5.28)

Table 2: Mean and Standard Deviation of $\ln(B_{31})$: $k = 5$, $\tilde{g}_2$=0

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	-2.12(1.27)	-2.31(1.07)	-2.44(1.14)	-2.61(0.93)	-3.37(2.71)
	IZS	-2.7(1.2)	-3.33(1.08)	-4(1.16)	-4.97(1.14)	-9.37(4.02)
	TESS	-1.81(1.09)	-2.39(0.99)	-3.02(1.08)	-3.76(0.83)	-4.49(1.02)
$p_2 = 5$	BG	-4.99(1.47)	-5.14(1.53)	-5.55(1.54)	-5.68(1.4)	-7.2(3.12)
	IZS	-5.61(1.38)	-6.59(1.44)	-7.87(1.48)	-9.43(1.54)	-14.85(4.38)
	TESS	-3.08(1.19)	-3.86(1.26)	-4.97(1.37)	-6.16(1.34)	-7.94(1.48)
$p_2 = 10$	BG	-12.28(2.48)	-13.25(2.56)	-14.6(2.23)	-15.44(2.57)	-18.9(3.91)
	IZS	-12.53(2.25)	-14.65(2.14)	-17.74(2.03)	-20.74(2.66)	-28.33(4.87)
	TESS	-4.93(1.68)	-6.31(1.79)	-8.58(1.8)	-10.96(2.19)	-15.31(2.11)
$p_2 = 20$	BG	-27.28(3.54)	-31.6(3.75)	-34.96(4.15)	-38.37(3.5)	-45.75(4.5)
	IZS	-27.75(3.46)	-32.42(3.2)	-38.27(3.81)	-45.33(3.66)	-57.39(6.11)
	TESS	-7.53(2.03)	-9.81(2.4)	-13.52(2.93)	-19.1(2.85)	-26.69(3.34)
$p_2 = 50$	BG	-80.8(7.94)	-92.13(6.83)	-105.24(6.96)	-118.65(6.87)	-137.51(8.39)
	IZS	-83.36(7.52)	-95.95(7.21)	-112.5(7.2)	-130.27(6.87)	-155.93(8.43)
	TESS	-12.23(3.14)	-17.52(3.81)	-27.92(5.01)	-39.02(6.09)	-57.74(6.91)

Table 3: Mean and Standard Deviation of $\ln(B_{31})$: $k = 20$, $\tilde{g}_2$=0

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	-3.55(0.98)	-3.52(1.22)	-3.86(1.03)	-3.8(0.89)	-4.57(2.75)
	IZS	-4.2(0.93)	-4.62(1.21)	-5.51(1)	-6.38(1.06)	-9.98(3.9)
	TESS	-3.16(0.91)	-3.56(1.19)	-4.43(0.97)	-5.05(0.86)	-5.88(1.11)
$p_2 = 5$	BG	-7.45(1.64)	-8.05(1.56)	-8.12(1.48)	-8.41(1.44)	-9.94(2.9)
	IZS	-8.28(1.58)	-9.34(1.44)	-10.46(1.44)	-12.19(1.48)	-16.91(4.48)
	TESS	-5.3(1.51)	-6.27(1.4)	-7.35(1.38)	-8.88(1.36)	-10.9(1.28)
$p_2 = 10$	BG	-18.07(2.67)	-19.67(2.27)	-20.71(2.3)	-21.57(2.48)	-24.52(3.79)
	IZS	-18.58(2.29)	-20.64(2.08)	-23.83(2.16)	-26.94(2.33)	-34.79(4.37)
	TESS	-9.27(2.08)	-11.02(1.9)	-13.96(2.05)	-16.71(2.32)	-20.94(2.38)
$p_2 = 20$	BG	-40.03(3.8)	-44.09(3.89)	-47.87(3.54)	-51.41(3.97)	-57.81(5.46)
	IZS	-40.51(3.86)	-45.06(3.35)	-51.7(3.37)	-58.78(3.75)	-70.95(4.99)
	TESS	-15.14(3.03)	-18.95(2.91)	-24.25(3.32)	-30.59(3.35)	-39.39(3.29)
$p_2 = 50$	BG	-113.86(7.58)	-126.34(7.35)	-139.44(5.96)	-151.78(6.09)	-172.38(7.59)
	IZS	-116.62(6.89)	-131.09(7.19)	-147.81(6.89)	-164(7)	-190.6(7.55)
	TESS	-30.53(4.76)	-39.55(5.96)	-54.58(6.11)	-70.31(6.8)	-91.7(6.47)

Table 4: Mean and Standard Deviation of $\ln(B_{31})$: $k = 3$, $\tilde{g}_2$=0.1

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	-0.87(1.63)	-0.36(2.53)	0.87(3.31)	3.36(6.08)	14.54(14.77)
	IZS	-1.44(1.68)	-1.26(2.58)	-0.59(3.43)	1.26(6.37)	8.86(14.66)
	TESS	-0.76(1.47)	-0.56(2.31)	0.09(3.11)	1.96(5.87)	11.61(13.41)
$p_2 = 5$	BG	-2.66(2.18)	-1.71(3.65)	1.23(5.03)	5.51(8.28)	21.89(19.43)
	IZS	-3.3(2.16)	-2.92(3.71)	-0.99(5.07)	2.16(8.48)	14.13(19.49)
	TESS	-1.31(1.76)	-0.8(3.2)	1.16(4.59)	4.44(7.79)	18.75(18.45)
$p_2 = 10$	BG	-6.97(3.81)	-5.35(4.82)	-0.16(7.06)	10.33(12.32)	48.27(27.31)
	IZS	-7.31(3.73)	-6.7(4.58)	-3.04(7.15)	5.02(12.4)	37.46(27.72)
	TESS	-1.49(2.84)	-0.36(3.7)	3.73(6.24)	12.2(11.4)	47.31(25.92)
$p_2 = 20$	BG	-15.11(6.61)	-13.86(7.53)	-4.82(11.15)	20.27(18.77)	97.39(40.55)
	IZS	-14.9(5.85)	-14.46(6.84)	-8.61(10.5)	12.82(18.93)	84.3(40.6)
	TESS	-0.31(3.87)	1.85(5)	9.43(8.89)	31.74(16.64)	106.02(38.27)
$p_2 = 50$	BG	-49.45(9.85)	-42.63(10.87)	-24.19(16.26)	37.55(30.37)	227.73(72.56)
	IZS	-51.55(10.69)	-47.51(12.12)	-32.3(17.47)	25.04(30.22)	209.53(72.12)
	TESS	1.3(5.1)	9.71(7.39)	31.08(13)	91.06(26.01)	280.07(67.34)

Table 5: Mean and Standard Deviation of $\ln(B_{31})$: $k = 5$, $\tilde{g}_2$=0.1

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	-0.89(2.13)	0.15(3.03)	1.53(4.72)	5.38(8.14)	23.41(22.5)
	IZS	-1.46(2.13)	-0.79(3.07)	0(4.84)	3.2(8.37)	17.27(22.88)
	TESS	-0.69(1.96)	-0.02(2.85)	0.78(4.58)	4.01(7.92)	20.39(21.67)
$p_2 = 5$	BG	-2.39(2.97)	-0.35(4.46)	2.88(6.28)	15.88(15.11)	38.29(37.34)
	IZS	-3.06(2.91)	-1.63(4.53)	0.59(6.33)	12.47(15.33)	29.79(37.58)
	TESS	-0.87(2.56)	0.61(4.13)	2.92(5.94)	14.63(14.6)	35.22(36.04)
$p_2 = 10$	BG	-5.35(5.11)	-4.25(6.43)	6.45(11.69)	24.32(20.04)	88.43(55.78)
	IZS	-5.91(4.96)	-5.39(6.27)	3.28(11.7)	18.99(20.09)	77.8(55.95)
	TESS	0.34(4.03)	1.51(5.37)	10.36(10.72)	26.41(19)	87.46(54.01)
$p_2 = 20$	BG	-14.69(7.51)	-8.61(9.36)	6.6(14.66)	52.44(31.49)	172.78(67.7)
	IZS	-13.91(7.19)	-9.88(9.25)	3.01(14.89)	44.94(31.89)	159.35(68.09)
	TESS	2.33(5.2)	7.73(7.2)	22.04(12.96)	64.32(29.62)	182.06(65.56)
$p_2 = 50$	BG	-45.22(11.82)	-35.66(15.57)	9.32(28.39)	112.31(46.98)	437.77(112.72)
	IZS	-49.12(12.57)	-39.57(15.78)	0.3(29.13)	100.14(47.22)	420.25(112.29)
	TESS	8.74(7.16)	22.66(11.31)	66.93(24.02)	167.9(43.21)	491.93(107.32)

Table 6: Mean and Standard Deviation of $\ln(B_{31})$: $k = 20$, $\tilde{g}_2$=0.1

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	2.31(6.63)	4.7(8.32)	12.39(15.44)	38.29(35.92)	91.07(96.48)
	IZS	1.71(6.7)	3.74(8.36)	10.8(15.54)	35.94(36.23)	84.77(96.68)
	TESS	2.58(6.52)	4.62(8.2)	11.64(15.31)	36.79(35.75)	88.1(95.85)
$p_2 = 5$	BG	3.07(8.28)	9.03(13.28)	32.22(26.82)	68.73(49.69)	172.2(123.37)
	IZS	2.54(8.25)	7.7(13.2)	29.95(26.96)	65.15(49.84)	163.76(124.26)
	TESS	5.02(7.91)	10.23(12.82)	32.31(26.49)	67.55(49.23)	169.14(122.42)
$p_2 = 10$	BG	4.83(11.81)	23.12(20.79)	58.92(43.33)	143.06(80.3)	388.47(184.21)
	IZS	4.53(11.81)	21.81(20.91)	55.73(43.41)	137.59(80.24)	378.58(184.74)
	TESS	12(10.94)	29.26(19.95)	63.35(42.35)	145.23(79.11)	388.34(183.25)
$p_2 = 20$	BG	11.6(18.07)	43.46(32.62)	127.12(55.66)	317.8(112.97)	826.28(278.66)
	IZS	11.05(18.25)	42.04(32.42)	123.3(55.49)	310.69(113.18)	815.39(279)
	TESS	30.3(16.32)	61.8(30.31)	143.25(53.53)	330.26(110.88)	836.96(276.06)
$p_2 = 50$	BG	27.88(29.23)	105.7(49.07)	341.01(95.71)	770.93(207.88)	2161.1(452.02)
	IZS	23.2(28.31)	100.89(49.07)	332.62(96.23)	760.61(207.77)	2144.63(451.8)
	TESS	89.87(24.52)	168.79(44.77)	401.13(91.94)	829.38(203.18)	2214.98(447.69)

Table 7: Mean and Standard Deviation of $\ln(B_{31})$: $k = 3$, $\tilde{g}_2$=0.3

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	0.1(2.68)	1.38(3.6)	5.54(6.92)	13.61(13.75)	38.76(36.7)
	IZS	-0.42(2.74)	0.54(3.73)	4.3(7.23)	12(14.41)	33.97(38.49)
	TESS	0.15(2.44)	1.05(3.36)	4.55(6.63)	11.85(13.31)	34.64(34.83)
$p_2 = 5$	BG	0.67(4.44)	4.35(6.16)	8.66(9.19)	25.91(21.64)	68.36(44.63)
	IZS	0(4.45)	3.17(6.19)	6.59(9.37)	23.07(22.2)	62.04(45.66)
	TESS	1.48(3.83)	4.53(5.52)	8.04(8.55)	23.96(20.84)	63.55(42.83)
$p_2 = 10$	BG	0.2(6.57)	6.15(9.44)	22.4(14.47)	54.02(29.4)	156.77(68.73)
	IZS	-0.38(6.35)	4.7(9.3)	19.56(14.54)	49.16(29.82)	147.49(69.41)
	TESS	3.92(5.14)	9.18(8.03)	23.92(13.24)	53.37(28.05)	152.13(66.6)
$p_2 = 20$	BG	-4.34(8.87)	10.74(12.61)	42.79(23.6)	116.87(50.21)	325.76(102.46)
	IZS	-3.97(8.41)	9.96(12.62)	39.09(23.71)	109.97(50.51)	313.8(101.75)
	TESS	7.31(6.26)	21.11(10.53)	50.95(21.27)	121.72(47.46)	325.95(98.33)
$p_2 = 50$	BG	-14.98(16.06)	17.54(20.96)	102.48(39.88)	268.2(68.44)	854.12(179.44)
	IZS	-17.48(15.55)	13.18(20.63)	94.86(39.71)	258.07(69.64)	836.89(178.1)
	TESS	22.45(10.72)	54.67(16.34)	138.18(35.23)	301.73(64.33)	881.77(173.24)

Table 8: Mean and Standard Deviation of $\ln(B_{31})$: $k = 5$, $\tilde{g}_2$=0.3

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	2.05(4.59)	4.24(6.49)	6.88(8.41)	22.64(22.8)	57.02(52.44)
	IZS	1.49(4.65)	3.37(6.61)	5.49(8.63)	20.96(23.49)	52.2(53.54)
	TESS	2.03(4.33)	3.86(6.2)	5.96(8.16)	20.92(22.39)	53.22(50.7)
$p_2 = 5$	BG	2.74(6.22)	6.96(9.1)	19.5(18.71)	51.93(35.57)	120.5(85.54)
	IZS	2.05(6.25)	5.69(9.22)	17.43(18.94)	49.17(36.25)	113.71(86.86)
	TESS	3.67(5.63)	7.31(8.55)	18.83(18.01)	49.89(34.9)	115.7(83.63)
$p_2 = 10$	BG	3.67(7.75)	13.72(11.77)	44.08(26.91)	94.62(47.24)	292.8(128.82)
	IZS	3.07(7.65)	12.41(11.85)	41.19(26.98)	89.61(47.53)	283.68(128.95)
	TESS	7.86(6.58)	17.26(10.67)	45.71(25.52)	93.97(46.05)	287.72(125.63)
$p_2 = 20$	BG	9.3(12.97)	32.11(20.39)	82.44(34.48)	208.26(71.9)	569.8(177.36)
	IZS	9.27(12.84)	30.91(20.25)	78.6(34.45)	201.47(72.14)	559.1(178.25)
	TESS	20.71(10.59)	42.64(18.08)	91.12(32.4)	213.43(69.56)	570.12(174.13)
$p_2 = 50$	BG	17.69(21.4)	68.57(29.74)	209.26(56.95)	522.87(112.7)	1456.2(286)
	IZS	14.49(21.59)	63.46(29.96)	201.86(57.84)	511.76(112.67)	1438.75(286.26)
	TESS	55.89(16.97)	107.4(25.43)	246.51(53.29)	556.17(108.64)	1485.18(281.32)

Table 9: Mean and Standard Deviation of $\ln(B_{31})$: $k = 20$, $\tilde{g}_2$=0.3

	Method	$p_1 = 3$	$p_1 = 5$	$p_1 = 10$	$p_1 = 20$	$p_1 = 50$
$p_2 = 3$	BG	12.29(14.54)	20.74(21.33)	48.39(51.63)	103.76(97.45)	248.89(204.9)
	IZS	11.73(14.57)	19.87(21.41)	47.05(52.01)	102.04(98.1)	243.67(205.8)
	TESS	12.37(14.27)	20.44(21.04)	47.38(51.34)	102.07(97.05)	245.11(203.1)
$p_2 = 5$	BG	23.92(17.61)	47.27(31.85)	87.35(51.58)	192.6(137.1)	457.19(313.5)
	IZS	23.22(17.54)	45.99(31.82)	85.17(51.78)	189.57(137.9)	451.74(314.5)
	TESS	24.97(17.02)	47.61(31.2)	86.7(51.08)	190.77(136.6)	452.8(311.6)
$p_2 = 10$	BG	54.14(33.18)	93.95(44.43)	219.38(98.92)	444.46(208.5)	1233.9(489.4)
	IZS	53.61(33.27)	92.47(44.47)	216.47(98.98)	439.6(208.7)	1224.9(489.6)
	TESS	58.6(32.01)	97.53(43.32)	221.15(97.52)	443.85(207.0)	1229.1(487.0)
$p_2 = 20$	BG	108.69(47.82)	205.48(83.05)	445.2(139.2)	907.64(261.3)	2391.0(714.9)
	IZS	108.22(48.3)	204.69(83.2)	441.8(138.7)	900.64(261.9)	2378.8(715.4)
	TESS	121.05(45.71)	217.47(80.47)	454.07(136.9)	912.93(259.2)	2391.5(711.8)
$p_2 = 50$	BG	263.2(73)	514.34(106.0)	1135.5(212.8)	2403.5(439.5)	6198.0(1029.7)
	IZS	261.92(73.38)	510.2(105.5)	1127.6(213.3)	2392.5(439.9)	6181.4(1030.0)
	TESS	306.24(68.37)	555.64(102.1)	1173.8(208.7)	2437.6(435.5)	6226.6(1025.0)