Supplementary Material for

Some Dimension Reduction Strategies for the Analysis of Survey Data

Jiaying Weng and Derek S. Young

Department of Statistics, University of Kentucky

October 29, 2017

1 Additional Tables and Figures

- Table 1 gives the variable numbers and definitions of variables used from the 2015 Planning Database.
- Tables 2 – 16 give the marginal tests for dimension using SIR, partial SIR, SAVE, yPHD, and rPHD, each for the responses Y_1, Y_2, and Y_3.
- Table 17 gives the coefficients for the six largest principal components. Tables 18 – 20 give the coefficients of the directions using SIR for each of the responses Y_1, Y_2, and Y_3.
- Figures 1 and 2 give the partial residual plots for the estimated additive models when Y_2 is the response. Figures 3 and 4 give the partial residual plots for the estimated additive models when Y_3 is the response.

No.	Definitions	Notation
73	Number of people ages 25 years and over at the time of interview with a college degree or higher in the ACS population	X_1
77	Number of people classified as below the poverty level given their total family income within the last year, family size, and family composition in the ACS population	X_2
103	Number of ACS households in which the householder and his or her spouse are listed as members of the same household; does not include same-sex married couples	X_3
112	Number of ACS households where a householder lives alone or with nonrelatives only; includes same-sex couples where no relatives of the householder are present	X_4
115	Number of ACS households where a householder lives alone	X_5
124	Number of ACS families with related children under 6 years old	X_6
130	Median ACS household income for the block group	X_7
132	Median ACS household income for the tract	X_8
145	Number of 2010 Census occupied housing units that are not owner occupied, whether they are rented or occupied without payment of rent	X_9
149	Number of ACS housing units where owner or co-owner lives in it	X_{10}
151	Number of ACS housing units in which the structure contains only that single unit	X_{11}
153	Number of ACS housing units in which the structure contains 2 or more housing units	X_{12}
155	Number of ACS housing units in which the structure contains 10 or more housing units	X_{13}
167	Median of ACS respondents' house value estimates for the block group	X_{14}
169	Median of ACS respondents' house value estimates for the tract	X_{15}
3	Name of State or statistically equivalent territory; island territories are excluded from this analysis; these values are converted to a categorical variable based on 9 Census-designated geographical regions; only used for partial SIR analysis	W
83	Number of people with no health insurance coverage in the ACS	Y_1
79	Number of people with one type of health insurance coverage in the ACS	Y_2
81	Number of people with two or more types of health insurance coverage in the ACS	Y_3

Table 1: The first two columns are the variable numbers and definitions as given in the 2015 Planning Database documentation. The last column is the variable notation used in our manuscript.

	Stat	df	p.value
0D vs >= 1D	127778.6966	1350.0000	0.0000
1D vs >= 2D	16024.8353	1246.0000	0.0000
2D vs >= 3D	3047.0326	1144.0000	0.0000
3D vs >= 4D	1734.3518	1044.0000	0.0000
4D vs >= 5D	1066.5385	946.0000	0.0037
5D vs >= 6D	887.9807	850.0000	0.1778

Table 2: Marginal tests for dimension using SIR for Y_1.

	Stat	df	p.value
0D vs >= 1D	190395.5193	1440.0000	0.0000
1D vs >= 2D	14939.5602	1330.0000	0.0000
2D vs >= 3D	3219.2447	1222.0000	0.0000
3D vs >= 4D	2105.2266	1116.0000	0.0000
4D vs >= 5D	1446.9417	1012.0000	0.0000
5D vs >= 6D	985.2887	910.0000	0.0415
6D vs >= 7D	819.7028	810.0000	0.3987

Table 3: Marginal tests for dimension using SIR for Y_2.

	Stat	df	p.value
0D vs >= 1D	122357.2786	1350.0000	0.0000
1D vs >= 2D	4365.2667	1246.0000	0.0000
2D vs >= 3D	1652.9354	1144.0000	0.0000
3D vs >= 4D	1090.7890	1044.0000	0.1530

Table 4: Marginal tests for dimension using SIR for Y_3.

	Stat	df	p.value
0D vs >= 1D	135883.3292	1890.0000	0.0000
1D vs >= 2D	30601.3194	1750.0000	0.0000
2D vs >= 3D	17235.6722	1612.0000	0.0000
3D vs >= 4D	11202.4369	1476.0000	0.0000
4D vs >= 5D	5729.4764	1342.0000	0.0000
5D vs >= 6D	4036.6166	1210.0000	0.0000
6D vs >= 7D	3074.6650	1080.0000	0.0000
7D vs >= 8D	2225.1084	952.0000	0.0000
8D vs >= 9D	1611.4217	826.0000	0.0000
9D vs >= 10D	1143.4266	702.0000	0.0000
10D vs >= 11D	838.7606	580.0000	0.0000
11D vs >= 12D	564.9019	460.0000	0.0006
12D vs >= 13D	378.5320	342.0000	0.0845

Table 5: Marginal tests for dimension using partial SIR for Y_1.

	Stat	df	p.value
0D vs >= 1D	186797.3133	1890.0000	0.0000
1D vs >= 2D	19837.5882	1750.0000	0.0000
2D vs >= 3D	11290.3427	1612.0000	0.0000
3D vs >= 4D	6349.3462	1476.0000	0.0000
4D vs >= 5D	4578.4808	1342.0000	0.0000
5D vs >= 6D	3574.0556	1210.0000	0.0000
6D vs >= 7D	2712.0028	1080.0000	0.0000
7D vs >= 8D	2101.2728	952.0000	0.0000
8D vs >= 9D	1544.1730	826.0000	0.0000
9D vs >= 10D	1090.4002	702.0000	0.0000
10D vs >= 11D	713.0341	580.0000	0.0001
11D vs >= 12D	520.1782	460.0000	0.0271
12D vs >= 13D	354.5500	342.0000	0.3087

Table 6: Marginal tests for dimension using partial SIR for Y_2.

	Stat	df	p.value
0D vs >= 1D	125037.5661	1890.0000	0.0000
1D vs >= 2D	12174.7443	1750.0000	0.0000
2D vs >= 3D	7708.3127	1612.0000	0.0000
3D vs >= 4D	5508.1238	1476.0000	0.0000
4D vs >= 5D	3601.2730	1342.0000	0.0000
5D vs >= 6D	2592.4299	1210.0000	0.0000
6D vs >= 7D	2036.2467	1080.0000	0.0000
7D vs >= 8D	1586.4564	952.0000	0.0000
8D vs >= 9D	1169.7110	826.0000	0.0000
9D vs >= 10D	883.9065	702.0000	0.0000
10D vs >= 11D	662.7294	580.0000	0.0096
11D vs >= 12D	503.9640	460.0000	0.0767

Table 7: Marginal tests for dimension using partial SIR for Y_3.

	Stat	df(Nor)	p.value(Nor)	p.value(Gen)
0D vs >= 1D	155893.0352	10800.0000	0.0000	0.0000
1D vs >= 2D	100362.0724	9450.0000	0.0000	0.0000
2D vs >= 3D	67439.6073	8190.0000	0.0000	0.0000
3D vs >= 4D	51215.0052	7020.0000	0.0000	0.0000
4D vs >= 5D	35688.0690	5940.0000	0.0000	0.0000
5D vs >= 6D	26905.6040	4950.0000	0.0000	0.0000
6D vs >= 7D	19995.1203	4050.0000	0.0000	0.0000
7D vs >= 8D	14602.3232	3240.0000	0.0000	0.0000
8D vs >= 9D	9537.5269	2520.0000	0.0000	0.0000
9D vs >= 10D	6363.6019	1890.0000	0.0000	0.0000
10D vs >= 11D	4106.7176	1350.0000	0.0000	0.0000
11D vs >= 12D	2501.2630	900.0000	0.0000	0.0000
12D vs >= 13D	963.0952	540.0000	0.0000	0.0000
13D vs >= 14D	481.3117	270.0000	0.0000	0.0000
14D vs >= 15D	165.6979	90.0000	0.0000	0.0001

Table 8: Marginal tests for dimension using SAVE for Y_1.

	Stat	df(Nor)	p.value(Nor)	p.value(Gen)
0D vs >= 1D	202380.8499	11520.0000	0.0000	0.0000
1D vs >= 2D	120104.8546	10080.0000	0.0000	0.0000
2D vs >= 3D	86294.5554	8736.0000	0.0000	0.0000
3D vs >= 4D	67227.5472	7488.0000	0.0000	0.0000
4D vs >= 5D	50647.6075	6336.0000	0.0000	0.0000
5D vs >= 6D	38796.3954	5280.0000	0.0000	0.0000
6D vs >= 7D	29546.9660	4320.0000	0.0000	0.0000
7D vs >= 8D	20973.2996	3456.0000	0.0000	0.0000
8D vs >= 9D	13229.5670	2688.0000	0.0000	0.0000
9D vs >= 10D	8881.6138	2016.0000	0.0000	0.0000
10D vs >= 11D	5079.2924	1440.0000	0.0000	0.0000
11D vs >= 12D	2796.0802	960.0000	0.0000	0.0000
12D vs >= 13D	1608.2308	576.0000	0.0000	0.0000
13D vs >= 14D	920.0632	288.0000	0.0000	0.0000
14D vs >= 15D	130.5440	96.0000	0.0110	0.2348

Table 9: Marginal tests for dimension using SAVE for Y_2.

	Stat	df(Nor)	p.value(Nor)	p.value(Gen)
0D vs >= 1D	165243.9364	10800.0000	0.0000	0.0000
1D vs >= 2D	93909.3811	9450.0000	0.0000	0.0000
2D vs >= 3D	72197.5033	8190.0000	0.0000	0.0000
3D vs >= 4D	53503.1774	7020.0000	0.0000	0.0000
4D vs >= 5D	36097.7707	5940.0000	0.0000	0.0000
5D vs >= 6D	30187.7898	4950.0000	0.0000	0.0000
6D vs >= 7D	21426.5150	4050.0000	0.0000	0.0000
7D vs >= 8D	14761.7767	3240.0000	0.0000	0.0000
8D vs >= 9D	11924.8792	2520.0000	0.0000	0.0000
9D vs >= 10D	8029.3853	1890.0000	0.0000	0.0000
10D vs >= 11D	4863.7098	1350.0000	0.0000	0.0000
11D vs >= 12D	2440.2639	900.0000	0.0000	0.0000
12D vs >= 13D	1374.6700	540.0000	0.0000	0.0000
13D vs >= 14D	660.7789	270.0000	0.0000	0.0000
14D vs >= 15D	263.5019	90.0000	0.0000	0.0000

Table 10: Marginal tests for dimension using SAVE for Y_3.

	Stat	df	p.value
0D vs >= 1D	177617.1046	120.0000	0.0000
1D vs >= 2D	88173.7126	105.0000	0.0000
2D vs >= 3D	45843.1292	91.0000	0.0000
3D vs >= 4D	29122.7571	78.0000	0.0000
4D vs >= 5D	19138.0857	66.0000	0.0000
5D vs >= 6D	13805.7601	55.0000	0.0000
6D vs >= 7D	9985.1348	45.0000	0.0000
7D vs >= 8D	6520.6016	36.0000	0.0000
8D vs >= 9D	3889.2340	28.0000	0.0000
9D vs >= 10D	1590.2069	21.0000	0.0000
10D vs >= 11D	942.9892	15.0000	0.0000
11D vs >= 12D	486.7179	10.0000	0.0000
12D vs >= 13D	71.1956	6.0000	0.0000
13D vs >= 14D	12.2295	3.0000	0.0066
14D vs >= 15D	1.0953	1.0000	0.2953

Table 11: Marginal tests for dimension using yPHD for Y_1.

	Stat	df	p.value
0D vs >= 1D	439738.0094	120.0000	0.0000
1D vs >= 2D	76715.4122	105.0000	0.0000
2D vs >= 3D	51136.6858	91.0000	0.0000
3D vs >= 4D	38958.6704	78.0000	0.0000
4D vs >= 5D	29988.2428	66.0000	0.0000
5D vs >= 6D	22442.1725	55.0000	0.0000
6D vs >= 7D	16056.8839	45.0000	0.0000
7D vs >= 8D	10865.5818	36.0000	0.0000
8D vs >= 9D	7673.7579	28.0000	0.0000
9D vs >= 10D	5090.7992	21.0000	0.0000
10D vs >= 11D	3086.2911	15.0000	0.0000
11D vs >= 12D	1695.3168	10.0000	0.0000
12D vs >= 13D	799.5905	6.0000	0.0000
13D vs >= 14D	157.2727	3.0000	0.0000
14D vs >= 15D	49.1525	1.0000	0.0000

Table 12: Marginal tests for dimension using yPHD for Y_2.

	Stat	df	p.value
0D vs >= 1D	353756.5493	120.0000	0.0000
1D vs >= 2D	66815.8666	105.0000	0.0000
2D vs >= 3D	42564.1906	91.0000	0.0000
3D vs >= 4D	29255.1540	78.0000	0.0000
4D vs >= 5D	21253.0088	66.0000	0.0000
5D vs >= 6D	15753.9311	55.0000	0.0000
6D vs >= 7D	11745.6460	45.0000	0.0000
7D vs >= 8D	7916.3094	36.0000	0.0000
8D vs >= 9D	5256.1625	28.0000	0.0000
9D vs >= 10D	3157.9192	21.0000	0.0000
10D vs >= 11D	1365.1721	15.0000	0.0000
11D vs >= 12D	377.5952	10.0000	0.0000
12D vs >= 13D	161.2063	6.0000	0.0000
13D vs >= 14D	23.7377	3.0000	0.0000
14D vs >= 15D	2.0988	1.0000	0.1474

Table 13: Marginal tests for dimension using yPHD for Y_3.

	Stat	df	Normal theory	General theory
0D vs >= 1D	63493.5068	120.0000	0.0000	0.0000
1D vs >= 2D	21521.6466	105.0000	0.0000	0.0000
2D vs >= 3D	14422.7581	91.0000	0.0000	0.0000
3D vs >= 4D	10123.7357	78.0000	0.0000	0.0000
4D vs >= 5D	7327.0031	66.0000	0.0000	0.0000
5D vs >= 6D	5386.9316	55.0000	0.0000	0.0000
6D vs >= 7D	3747.6975	45.0000	0.0000	0.0000
7D vs >= 8D	2532.2495	36.0000	0.0000	0.0000
8D vs >= 9D	1508.2198	28.0000	0.0000	0.0000
9D vs >= 10D	894.7523	21.0000	0.0000	0.0000
10D vs >= 11D	466.2296	15.0000	0.0000	0.0046
11D vs >= 12D	245.7336	10.0000	0.0000	0.0261
12D vs >= 13D	113.4531	6.0000	0.0000	0.1162

Table 14: Marginal tests for dimension using rPHD for Y_1.

	Stat	df	Normal theory	General theory
0D vs >= 1D	297042.0875	120.0000	0.0000	0.0000
1D vs >= 2D	57454.3761	105.0000	0.0000	0.0000
2D vs >= 3D	32396.9258	91.0000	0.0000	0.0000
3D vs >= 4D	21121.3081	78.0000	0.0000	0.0000
4D vs >= 5D	15731.3264	66.0000	0.0000	0.0000
5D vs >= 6D	11043.5915	55.0000	0.0000	0.0000
6D vs >= 7D	8151.5707	45.0000	0.0000	0.0000
7D vs >= 8D	6093.5342	36.0000	0.0000	0.0000
8D vs >= 9D	4369.9673	28.0000	0.0000	0.0000
9D vs >= 10D	3045.2298	21.0000	0.0000	0.0000
10D vs >= 11D	1774.6722	15.0000	0.0000	0.0000
11D vs >= 12D	981.6630	10.0000	0.0000	0.0000
12D vs >= 13D	448.7045	6.0000	0.0000	0.0000
13D vs >= 14D	202.5587	3.0000	0.0000	0.0000
14D vs >= 15D	32.6826	1.0000	0.0000	0.0000

Table 15: Marginal tests for dimension using rPHD for Y_2.

	Stat	df	Normal theory	General theory
0D vs >= 1D	186134.5644	120.0000	0.0000	0.0791
1D vs >= 2D	21586.0017	105.0000	0.0000	0.0000
2D vs >= 3D	12164.9782	91.0000	0.0000	0.0000
3D vs >= 4D	7662.8802	78.0000	0.0000	0.0000
4D vs >= 5D	5390.0871	66.0000	0.0000	0.0000
5D vs >= 6D	3877.5847	55.0000	0.0000	0.0000
6D vs >= 7D	2616.0759	45.0000	0.0000	0.0000
7D vs >= 8D	1464.6075	36.0000	0.0000	0.0000
8D vs >= 9D	723.0290	28.0000	0.0000	0.0000
9D vs >= 10D	283.0354	21.0000	0.0000	0.0000
10D vs >= 11D	124.2678	15.0000	0.0000	0.0000
11D vs >= 12D	64.9884	10.0000	0.0000	0.0001
12D vs >= 13D	31.5987	6.0000	0.0000	0.0153
13D vs >= 14D	12.3845	3.0000	0.0062	0.0721
14D vs >= 15D	1.7082	1.0000	0.1912	0.3870

Table 16: Marginal tests for dimension using rPHD for Y_3.

	PC1	PC2	PC3	PC4	PC5	PC6
X_1	0.3525	-0.2220	-0.0804	0.0792	-0.0381	0.0175
X_2	-0.2199	-0.2521	0.1719	-0.4050	0.5662	0.0452
X_3	0.3206	-0.1968	0.2861	-0.1045	-0.0947	0.0107
X_4	-0.0243	-0.4363	-0.0496	0.3840	0.1575	0.0010
X_5	-0.0285	-0.4231	-0.0289	0.4330	0.1240	0.0186
X_6	0.0926	-0.2500	0.2436	-0.5910	-0.2395	0.1838
X_7	0.3995	0.0907	-0.1207	-0.0124	-0.2023	0.0353
X_8	0.3909	0.0485	-0.1644	-0.0141	-0.0959	0.0623
X_9	-0.1554	-0.4076	-0.1521	-0.1533	0.0054	0.0266
X_{10}	0.3047	-0.1725	0.3625	0.1379	-0.0128	-0.0539
X_{11}	0.2492	-0.1457	0.4150	0.0522	0.0625	-0.2072
X_{12}	-0.1501	-0.2912	-0.2521	-0.1776	-0.3920	-0.7522
X_{13}	-0.0893	-0.3204	-0.2985	-0.0264	-0.3511	0.5796
X_{14}	0.3141	-0.0186	-0.3770	-0.1696	0.3445	-0.0818
X_{15}	0.3065	-0.0320	-0.3955	-0.1672	0.3487	-0.0581

Table 17: Coefficients for the six largest principal components.

	Dir1	Dir2	Dir3	Dir4	Dir5
X_1	0.2073	0.1264	0.1019	0.1734	0.0052
X_2	-0.3213	-0.2656	-0.1104	-0.0487	0.2663
X_3	-0.0989	-0.1650	0.1767	-0.1665	-0.0581
X_4	-0.1878	0.2178	0.1142	-0.0637	-0.4361
X_5	0.3161	-0.2274	-0.1972	0.0662	0.2631
X_6	-0.1585	-0.2003	-0.0050	-0.0779	-0.2062
X_7	-0.0241	-0.7045	-0.2246	0.0280	0.0450
X_8	0.3478	-0.1335	0.5078	-0.6525	0.3294
X_9	-0.5766	0.2570	-0.1990	0.3248	-0.1637
X_{10}	-0.0449	0.1591	-0.0820	0.2510	-0.1134
X_{11}	-0.0122	0.0328	0.0063	0.0227	0.1557
X_{12}	0.1381	0.0358	0.1548	-0.1365	0.2951
X_{13}	0.0854	-0.3300	0.1676	-0.3942	0.1226
X_{14}	0.1468	-0.1690	-0.0940	0.1100	0.2672
X_{15}	-0.4238	-0.0403	-0.6918	0.3791	-0.5253

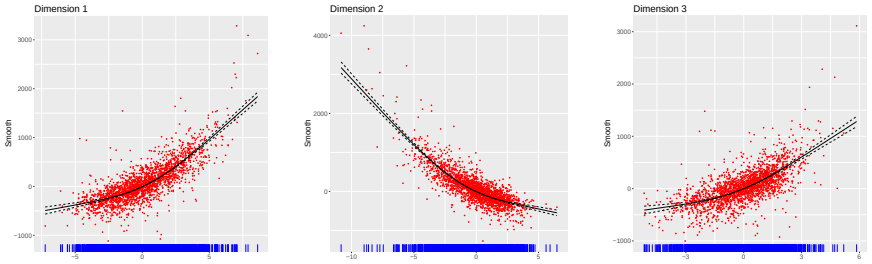
Table 18: Coefficients of directions using SIR for Y_1.

	Dir1	Dir2	Dir3	Dir4	Dir5	Dir6
X_1	-0.2204	-0.2731	-0.1013	0.1569	-0.1063	-0.1416
X_2	-0.2516	0.3307	0.0180	0.0317	0.0848	0.0342
X_3	-0.2335	-0.0153	-0.0152	0.0681	0.0184	0.0329
X_4	-0.2898	0.0513	0.3301	-0.0985	-0.0723	-0.1500
X_5	0.3778	0.2154	-0.0646	0.0549	0.1543	0.3391
X_6	-0.4415	0.0471	0.2814	-0.1028	-0.1108	-0.0406
X_7	-0.2766	0.7754	0.3332	0.0188	0.0740	0.2524
X_8	-0.0906	0.0263	-0.0904	0.6366	0.0288	0.1055
X_9	-0.5272	-0.1434	-0.3863	0.1032	0.6010	-0.0869
X_{10}	-0.1943	0.0430	-0.1205	-0.0472	0.0404	-0.0954
X_{11}	-0.0125	-0.1097	0.0251	-0.0308	-0.0296	0.0491
X_{12}	-0.0187	0.0879	0.0058	-0.0402	-0.1337	-0.0722
X_{13}	0.0589	-0.2055	-0.3231	-0.1296	-0.5565	0.1784
X_{14}	-0.0242	0.2015	0.0220	-0.0459	0.0207	-0.3383
X_{15}	-0.0842	-0.1943	-0.6390	-0.7117	-0.4918	0.7702

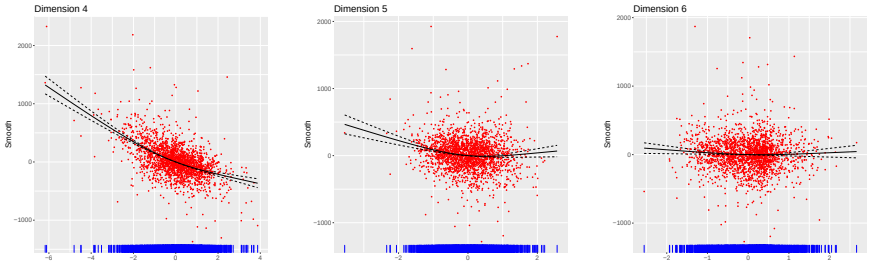
Table 19: Coefficients of directions using SIR for Y_2.

	Dir1	Dir2	Dir3
X_1	0.0807	0.0931	0.0374
X_2	-0.0615	-0.0727	0.0913
X_3	-0.3158	-0.0160	-0.1243
X_4	0.2033	0.1231	-0.2009
X_5	-0.4741	-0.2855	0.1517
X_6	0.2626	-0.0074	0.1238
X_7	0.3778	-0.0886	0.1529
X_8	0.4416	-0.5953	0.5560
X_9	-0.1200	0.2260	0.2358
X_{10}	-0.3198	0.0175	0.1309
X_{11}	-0.0346	0.0168	-0.0675
X_{12}	-0.0263	0.0510	-0.0105
X_{13}	-0.1888	0.0653	0.0046
X_{14}	-0.1453	-0.1032	0.1437
X_{15}	-0.2048	0.6766	-0.6823

Table 20: Coefficients of directions using SIR for Y_3.

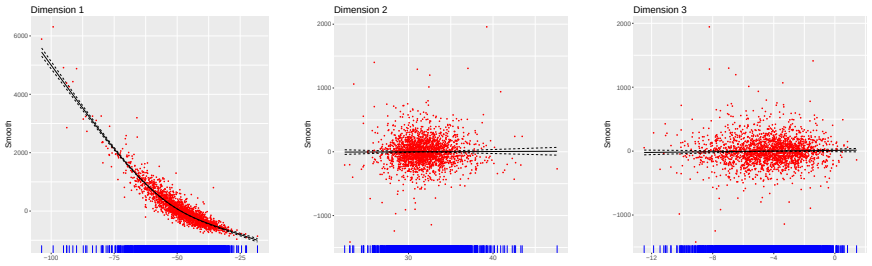


(a) (b) (c)

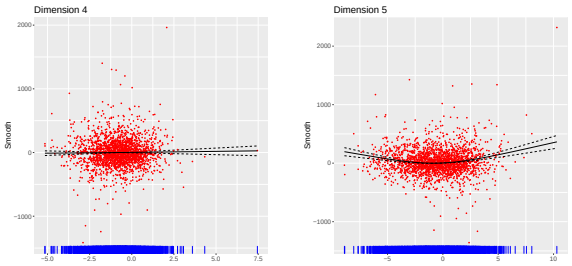


(d) (e) (f)

Figure 1: Partial residual plots for each of the PCA predictors when Y_2 is the response.

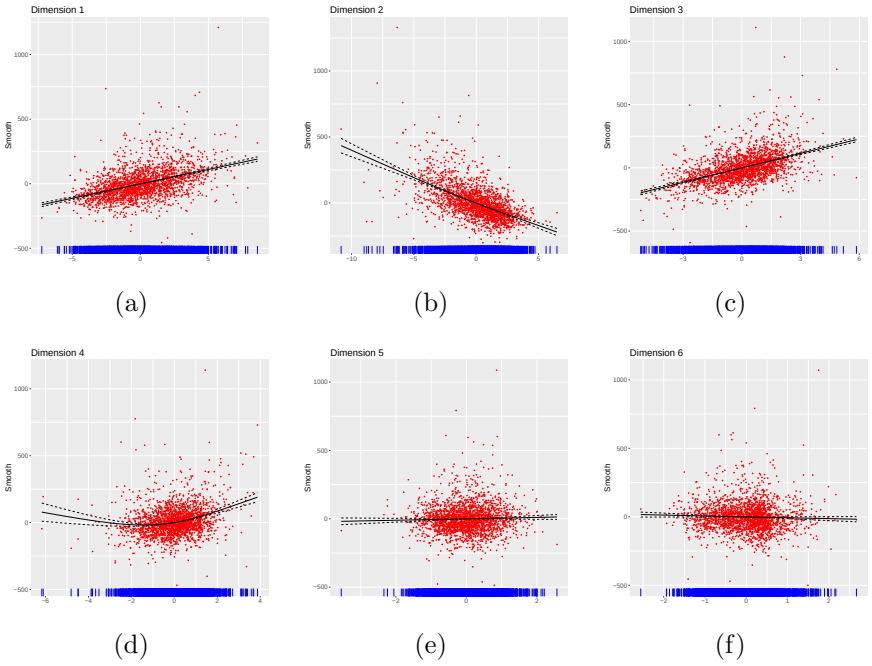


(a) (b) (c)



(d) (e)

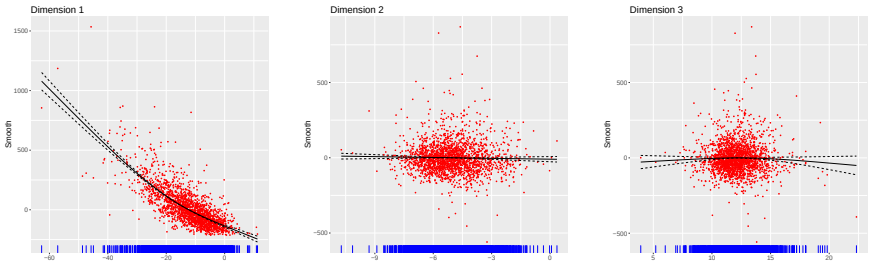
Figure 2: Partial residual plots for each of the significant SIR predictors when Y_2 is the response.



(a) (b) (c)

(d) (e) (f)

Figure 3: Partial residual plots for each of the PCA predictors when Y_3 is the response.



(a) (b) (c)

Figure 4: Partial residual plots for each of the significant SIR predictors when Y_3 is the response.