1

Worked Example for All Possible Regressions

This example illustrates the use of all possible regressions to select a subset of predictors to focus on under conditions of high dimensionality. All data are hypothetical. I assume you have read the material in the main text on all possible regressions. I will illustrate the variable selection approach of Cai, Tsay and Chen (2009) described in Chapter 11. I generated hypothetical data for 63 potential predictors of an outcome variable, Y, in which the predictors were intercorrelated with one another at a magnitude of about 0.20. In the spirit of Cai et al., I initially randomly divided the predictors into 9 classes of 7 predictors each. The first set of predictors is labeled A1 through A7, the second set B1 through B7, the third set C1 through C7, and so on through I1 to I7. The sample size was 250, which is small for 63 predictors (only 4 cases per predictor). There are 2,016 different correlations in the lower triangle of the correlation matrix between all variables! I approached the classification of predictors into the sets in a completely atheoretical fashion, i.e., as if set assignment was random. In practice, theory might help assist in the classification process, but I operate throughout the exercise as if I am theoretically blind. For convenience, all predictors were simulated to be normally distributed with a mean of zero and a standard deviation of 1.0.

Of the 63 predictors, only seven of them were defined so as to influence Y in accord with the following population regression equation Y = 0.40*A1 + 0.40*A2 + 0.40*C3 + 0.40*C4 + 0.40*D5 + 0.40*D6 + 0.40*E7. Thus, in reality, only variables A1, A2, C3, C4, D5, D6 and E7 impact Y. The overall squared R in the population for these seven predictors was set to 0.50 through the addition of normally distributed error. The influence of each of the seven predictors is moderate in magnitude, additive, and linear. Of interest is whether these seven variables are in the final set of predictors identified by the Cai et al. method as being relevant and how many false positive predictors enter the mix as well.

THE ANALYSIS

I used the program for all possible regressions on my website. As a first step, I specified Y as the outcome and the first set of variables A1 through A7 as predictors. The program begins by renaming the predictors and then specifying the regression equations that were evaluated using those new names, numbering the equations from m1 to m127 (the intercept is represented by the number 1, but it is formally estimated in each model):

```
PREDICTOR VARIABLES

x1: A1
x2: A2
x3: A3
x4: A4
x5: A5
x6: A6
x7: A7

MODELS EVALUATED (1 = Intercept)

m1: y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7
m2: y ~ 1 + x2 + x3 + x4 + x5 + x6 + x7
m3: y ~ 1 + x1 + x3 + x4 + x5 + x6 + x7
m4: y ~ 1 + x3 + x4 + x5 + x6 + x7
m5: y ~ 1 + x1 + x2 + x4 + x5 + x6 + x7
m6: y ~ 1 + x2 + x4 + x5 + x6 + x7
m7: y ~ 1 + x1 + x4 + x5 + x6 + x7
m8: y ~ 1 + x4 + x5 + x6 + x7
m9: y ~ 1 + x1 + x2 + x3 + x5 + x6 + x7
m10: y ~ 1 + x2 + x3 + x5 + x6 + x7
m11: y ~ 1 + x1 + x3 + x5 + x6 + x7
m12: y ~ 1 + x3 + x5 + x6 + x7
m13: y ~ 1 + x1 + x2 + x5 + x6 + x7
m14: y ~ 1 + x2 + x5 + x6 + x7
m15: y ~ 1 + x1 + x5 + x6 + x7
m16: y ~ 1 + x5 + x6 + x7
m17: y ~ 1 + x1 + x2 + x3 + x4 + x6 + x7
m18: y ~ 1 + x2 + x3 + x4 + x6 + x7
m19: y ~ 1 + x1 + x3 + x4 + x6 + x7
m20: y ~ 1 + x3 + x4 + x6 + x7
m21: y ~ 1 + x1 + x2 + x4 + x6 + x7
m22: y ~ 1 + x2 + x4 + x6 + x7
m23: y ~ 1 + x1 + x4 + x6 + x7
m24: y ~ 1 + x4 + x6 + x7
m25: y ~ 1 + x1 + x2 + x3 + x6 + x7
m26: y ~ 1 + x2 + x3 + x6 + x7
m27: y ~ 1 + x1 + x3 + x6 + x7
m28: y ~ 1 + x3 + x6 + x7
m29: y ~ 1 + x1 + x2 + x6 + x7
m30: y ~ 1 + x2 + x6 + x7
m31: y ~ 1 + x1 + x6 + x7
m32: y ~ 1 + x6 + x7
m33: y ~ 1 + x1 + x2 + x3 + x4 + x5 + x7
m34: y ~ 1 + x2 + x3 + x4 + x5 + x7
m35: y ~ 1 + x1 + x3 + x4 + x5 + x7
```

```
m36: y ~ 1 + x3 + x4 + x5 + x7
m37: y ~ 1 + x1 + x2 + x4 + x5 + x7
m38: y ~ 1 + x2 + x4 + x5 + x7
m39: y ~ 1 + x1 + x4 + x5 + x7
m40: y ~ 1 + x4 + x5 + x7
m41: y ~ 1 + x1 + x2 + x3 + x5 + x7
m42: y ~ 1 + x2 + x3 + x5 + x7
m43: y ~ 1 + x1 + x3 + x5 + x7
m44: y ~ 1 + x3 + x5 + x7
m45: y ~ 1 + x1 + x2 + x5 + x7
m46: y ~ 1 + x2 + x5 + x7
m47: y ~ 1 + x1 + x5 + x7
m48: y ~ 1 + x5 + x7
m49: y ~ 1 + x1 + x2 + x3 + x4 + x7
m50: y ~ 1 + x2 + x3 + x4 + x7
m51: y ~ 1 + x1 + x3 + x4 + x7
m52: y ~ 1 + x3 + x4 + x7
m53: y ~ 1 + x1 + x2 + x4 + x7
m54: y ~ 1 + x2 + x4 + x7
m55: y ~ 1 + x1 + x4 + x7
m56: y ~ 1 + x4 + x7
m57: y ~ 1 + x1 + x2 + x3 + x7
m58: y ~ 1 + x2 + x3 + x7
m59: y ~ 1 + x1 + x3 + x7
m60: y ~ 1 + x3 + x7
m61: y ~ 1 + x1 + x2 + x7
m62: y ~ 1 + x2 + x7
m63: y ~ 1 + x1 + x7
m64: y ~ 1 + x7
m65: y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6
m66: y ~ 1 + x2 + x3 + x4 + x5 + x6
m67: y ~ 1 + x1 + x3 + x4 + x5 + x6
m68: y ~ 1 + x3 + x4 + x5 + x6
m69: y ~ 1 + x1 + x2 + x4 + x5 + x6
m70: y ~ 1 + x2 + x4 + x5 + x6
m71: y ~ 1 + x1 + x4 + x5 + x6
m72: y ~ 1 + x4 + x5 + x6
m73: y ~ 1 + x1 + x2 + x3 + x5 + x6
m74: y ~ 1 + x2 + x3 + x5 + x6
m75: y ~ 1 + x1 + x3 + x5 + x6
m76: y ~ 1 + x3 + x5 + x6
m77: y ~ 1 + x1 + x2 + x5 + x6
m78: y ~ 1 + x2 + x5 + x6
m79: y ~ 1 + x1 + x5 + x6
m80: y ~ 1 + x5 + x6
m81: y ~ 1 + x1 + x2 + x3 + x4 + x6
m82: y ~ 1 + x2 + x3 + x4 + x6
m83: y ~ 1 + x1 + x3 + x4 + x6
m84: y ~ 1 + x3 + x4 + x6
```

```
m85: y ~ 1 + x1 + x2 + x4 + x6
m86: y ~ 1 + x2 + x4 + x6
m87: y ~ 1 + x1 + x4 + x6
m88: y ~ 1 + x4 + x6
m89: y ~ 1 + x1 + x2 + x3 + x6
m90: y ~ 1 + x2 + x3 + x6
m91: y ~ 1 + x1 + x3 + x6
m92: y ~ 1 + x3 + x6
m93: y ~ 1 + x1 + x2 + x6
m94: y ~ 1 + x2 + x6
m95: y ~ 1 + x1 + x6
m96: y ~ 1 + x6
m97: y ~ 1 + x1 + x2 + x3 + x4 + x5
m98: y ~ 1 + x2 + x3 + x4 + x5
m99: y ~ 1 + x1 + x3 + x4 + x5
m100: y ~ 1 + x3 + x4 + x5
m101: y ~ 1 + x1 + x2 + x4 + x5
m102: y ~ 1 + x2 + x4 + x5
m103: y ~ 1 + x1 + x4 + x5
m104: y ~ 1 + x4 + x5
m105: y ~ 1 + x1 + x2 + x3 + x5
m106: y ~ 1 + x2 + x3 + x5
m107: y ~ 1 + x1 + x3 + x5
m108: y ~ 1 + x3 + x5
m109: y ~ 1 + x1 + x2 + x5
m110: y ~ 1 + x2 + x5
m111: y ~ 1 + x1 + x5
m112: y ~ 1 + x5
m113: y ~ 1 + x1 + x2 + x3 + x4
m114: y ~ 1 + x2 + x3 + x4
m115: y ~ 1 + x1 + x3 + x4
m116: y ~ 1 + x3 + x4
m117: y ~ 1 + x1 + x2 + x4
m118: y ~ 1 + x2 + x4
m119: y ~ 1 + x1 + x4
m120: y ~ 1 + x4
m121: y ~ 1 + x1 + x2 + x3
m122: y ~ 1 + x2 + x3
m123: y ~ 1 + x1 + x3
m124: y ~ 1 + x3
m125: y ~ 1 + x1 + x2
m126: y ~ 1 + x2
m127: y ~ 1 + x1
```

Next, the results of each model are reported:

```
RESULTS

MODEL    AdjRSQR      AIC            BIC
```

1	0.311	1003.6	1035.3
2	0.252	1023.3	1051.4
3	0.252	1023.3	1051.4
4	0.189	1042.5	1067.2
5	0.307	1004.1	1032.2
6	0.235	1027.9	1052.6
7	0.242	1025.7	1050.3
8	0.160	1050.3	1071.4
9	0.311	1002.7	1030.9
10	0.248	1023.6	1048.3
11	0.247	1024.1	1048.8
12	0.177	1045.3	1066.4
13	0.306	1003.5	1028.1
14	0.228	1029.2	1050.3
15	0.234	1027.3	1048.4
16	0.142	1054.8	1072.4
17	0.312	1002.3	1030.5
18	0.251	1022.8	1047.5
19	0.247	1023.9	1048.6
20	0.179	1044.7	1065.8
21	0.308	1003.0	1027.6
22	0.231	1028.2	1049.3
23	0.235	1027.1	1048.3
24	0.144	1054.2	1071.8
25	0.312	1001.5	1026.2
26	0.246	1023.5	1044.7
27	0.240	1025.4	1046.5
28	0.163	1048.5	1066.1
29	0.306	1002.6	1023.7
30	0.222	1030.1	1047.7
31	0.224	1029.6	1047.2
32	0.118	1060.6	1074.6
33	0.307	1004.1	1032.3
34	0.243	1025.5	1050.2
35	0.241	1026.1	1050.7
36	0.169	1047.6	1068.8
37	0.303	1004.8	1029.4
38	0.223	1030.9	1052.0
39	0.229	1029.0	1050.2
40	0.135	1056.8	1074.4
41	0.306	1003.8	1028.4
42	0.235	1027.0	1048.1
43	0.231	1028.4	1049.5
44	0.149	1052.8	1070.4
45	0.300	1004.9	1026.0
46	0.211	1033.7	1051.3
47	0.215	1032.4	1050.0
48	0.103	1064.7	1078.8

49	0.307	1003.2	1027.8
50	0.239	1025.8	1047.0
51	0.232	1028.0	1049.1
52	0.152	1051.8	1069.4
53	0.302	1004.2	1025.3
54	0.216	1032.2	1049.8
55	0.216	1032.0	1049.7
56	0.107	1063.6	1077.7
57	0.305	1003.1	1024.3
58	0.229	1028.0	1045.6
59	0.218	1031.4	1049.0
60	0.123	1059.1	1073.2
61	0.298	1004.7	1022.3
62	0.200	1036.2	1050.2
63	0.197	1037.1	1051.2
64	0.061	1075.1	1085.7
65	0.302	1006.1	1034.3
66	0.244	1025.0	1049.7
67	0.235	1028.1	1052.8
68	0.173	1046.5	1067.6
69	0.299	1006.3	1030.9
70	0.228	1029.1	1050.3
71	0.225	1030.1	1051.3
72	0.146	1053.5	1071.1
73	0.302	1005.1	1029.7
74	0.241	1025.2	1046.3
75	0.230	1028.8	1049.9
76	0.162	1048.9	1066.5
77	0.298	1005.5	1026.6
78	0.222	1030.1	1047.7
79	0.218	1031.5	1049.1
80	0.128	1057.7	1071.8
81	0.300	1005.6	1030.3
82	0.239	1025.6	1046.8
83	0.223	1031.0	1052.1
84	0.154	1051.1	1068.7
85	0.296	1006.1	1027.2
86	0.221	1030.6	1048.2
87	0.210	1034.0	1051.6
88	0.119	1060.2	1074.3
89	0.300	1004.8	1025.9
90	0.235	1026.2	1043.8
91	0.215	1032.4	1050.0
92	0.138	1054.8	1068.9
93	0.295	1005.6	1023.2
94	0.213	1032.2	1046.3
95	0.199	1036.4	1050.5
96	0.094	1066.4	1077.0
97	0.292	1008.5	1033.2

98	0.227	1029.5	1050.6
99	0.212	1034.5	1055.6
100	0.139	1055.5	1073.1
101	0.288	1008.9	1030.0
102	0.209	1034.3	1051.9
103	0.200	1037.1	1054.7
104	0.106	1064.1	1078.2
105	0.290	1008.2	1029.3
106	0.220	1030.9	1048.5
107	0.201	1037.0	1054.6
108	0.117	1060.9	1075.0
109	0.286	1008.9	1026.5
110	0.198	1036.9	1051.0
111	0.186	1040.7	1054.7
112	0.072	1072.2	1082.8
113	0.288	1009.2	1030.3
114	0.217	1031.9	1049.5
115	0.189	1040.7	1058.3
116	0.103	1064.8	1078.9
117	0.283	1010.0	1027.6
118	0.194	1037.9	1052.0
119	0.172	1044.8	1058.9
120	0.055	1076.8	1087.4
121	0.285	1009.2	1026.8
122	0.206	1034.2	1048.3
123	0.171	1045.0	1059.1
124	0.068	1073.4	1084.0
125	0.278	1010.5	1024.5
126	0.178	1042.0	1052.6
127	0.148	1050.9	1061.5

I illustrate the approach to model selection that uses the BIC. First, I locate the model with the lowest BIC, which is model 61 with a BIC of 1022.3. This is the initial "best" model. The predictors in this model are A1, A2 and A7. Raftery (1995) suggests that if two models are within 2.2 BIC units of each other, they have about equal support. I therefore scan the list to determine if there is a more parsimonious model that is within 2.2 BIC units of 1022.3. Model 125 has 2 predictors, A1 and A2, and its BIC is within 2.2 units of model 61. I therefore used this model, selecting A1 and A2 from the first set as the "winning" predictors for my next iteration, described later.

I repeated this process for each of the remaining 8 sets of predictors, one set at a time. In cases where the number of predictors was the same in the most parsimonious equations that were within 2.2 BIC units of the lowest BIC value, I selected the equation with the lowest BIC. The unique predictors that entered the final winning pool after all nine sets were analyzed were A1, A2, B4, B5, B7, C3, C4, C6, D3, D5, D6, E1, E2, E7, F2, G1, H2 and I7, or 18 predictors. I randomly assigned these predictors to three sets of

6 predictors each and repeated the BIC model selection process for each of these three sets, separately, just as I did in the first iteration. The winning predictors from these three analyses were A1, A2, C3, C4, D5, D6, D7, E2 and E7, or only 9 predictors. I decided to submit all 9 predictors to an all possible regression analysis, with the final selected model that had the lowest BIC having the predictors A1, A2, C3, C4, D5, D6 and E7. These are the seven predictors that were indeed the population determinants of Y! One might then consider theory building around these seven variables.

Although the approach was effective in this case, there undoubtedly will be scenarios where it includes false positives or false negatives. Again, I caution against approaching veriable selection in a purely theory blind fashion.

REFERENCES

Cai, A., Tsay, R. & Chen, R. (2009). Variable selection in linear regression with many predictors, *Journal of Computational and Graphical Statistics*, 18, 573-591.

Raftery, A.E. (1995). Bayesian model selection in social research (with discussion). *Sociological Methodology*, 25, 111-195.