Correlation

February 24, 2017

1 DGP

Model 1

$$y = 3x_1^2x_5^2 + x_3 + 3.5x_4^2 + 3.2x_1 + 7x_2^3 + \varepsilon,$$

$(x_1, \ldots, x_5) \overset{D}{\sim} \mathcal{N}(0, I_5)$. $x_j = d_j + q\left(\sum_{i=1}^{5} x_i\right)$, d_j's are i.i.d. $\mathcal{N}(0,1)$, $p \geq j > 5$.

Model 2

$$y = 3.1x_1x_2 + 4.1\sin(x_3x_4) - 3.7x_5 - 4.2x_2 + \varepsilon,$$

$(x_1, \ldots, x_5) \overset{D}{\sim} \mathcal{N}(0, I_5)$. $x_j = d_j + q\left(\sum_{i=1}^{5} x_i\right)$, d_j's are i.i.d. $\mathcal{N}(0, \sqrt{0.25})$, $p \geq j > 5$.

Results

$n = 600$, $R = 30$, $p = 1000$, $q = (3/20)^{1/2}$, all relevent variables were retrieved in every case.

2 Algorithm

step1

Set $K_n = 10$. Grouped OGA to order 4 on y; expanding $x_{\hat{\jmath}_{1i}}, i = 1, \ldots, K_n$ to fourth order (polynomial); run the OGA + HDIC + Trim on y and this expended data set and get subset $\mathcal{A}_1$ as well as the corresponding residual $\hat{y}_1$.

step2

Find the grouped OGA path of length K_n for $\hat{y}_1$, denoting them as $x_{\hat{j}_{2i}}, i = 1, \ldots, K_n$; union these K_n variables and $\mathcal{A}_1$ and expand them to order 4 polynomials; run the OGA + HDIC + Trim on y(note this) and get subset $\mathcal{A}_2$ as well as the corresponding residual $\hat{y}_2$.

step3

Ideally, we need to find a rule for stopping these procedure so we can move on to y^2. Let's pretend we did it.

step4

Find the grouped OGA path of length K_n for $\hat{y}_2^2$, denoting them as $x_{\hat{j}_{3i}}, i = 1, \ldots, K_n$; union these K_n variables and $\mathcal{A}_2$ and expand them to order 4 polynomials; run the OGA + HDIC + Trim on y(note this) and get subset $\mathcal{A}_3$ as well as the corresponding residual $\hat{y}_3$.

step5

Again, here we need to find another rule to stop the procedure for y^2, and this should be the end of the while procedure. The last $\mathcal{A}_3$ is reported.

Remark. *In our cases, all relevent varialbes are included and no further gain in fitting power for antoher round. Step 2 is reuqired, however.*