Homework 1. Model selection in sparsity

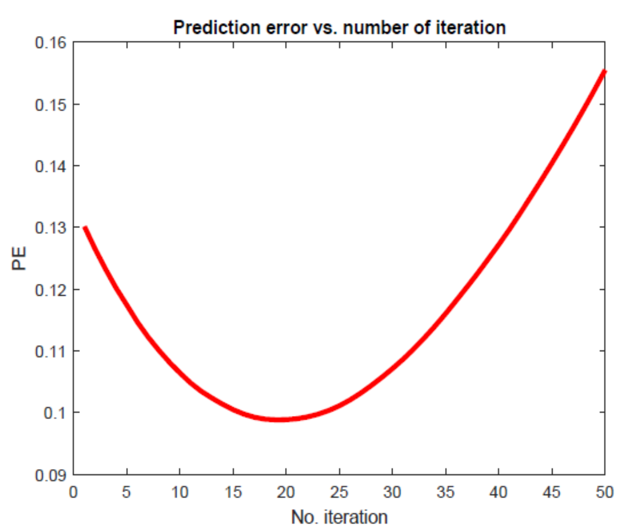
Student: Hongxing NIU;   Matricule ULB: 000342366

1. Simulate the observational model $Y = X\beta + \varepsilon$ where β is sparse. This is simulated by letting β be a realization of a random variable V which has probability 1 - p of being zero.

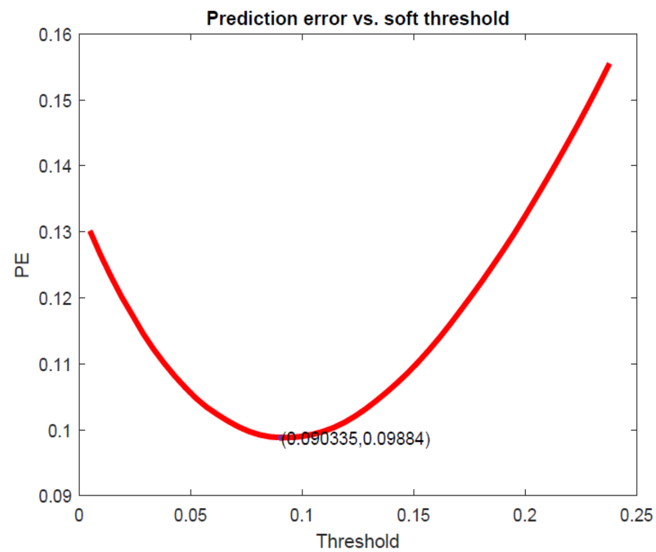
```
>> simulatesparseregressionSTATF408
-------------------------------------------------
Generating Y = X*beta+stdev*Z with following values:
   typeofdesign = random
             (columns of design (X) normalized, using ell-NaN-norm)
         n = 200
         m = 1000
         p = 0.05
       snr = 10
realrandom = 0
-------------------------------------------------


Simulation setup finished
Warning: iterativeST: maximum number of iterations reached
> In iterativeST (line 64)
  In simulatesparseregressionSTATF408 (line 28)
```

Where p = degree of sparsity (in zero-inflated Laplacian model of beta); *SNR (signal-to-noise ratio)*= $10 \times ln(\frac{\|X\beta\|_2^2}{\sigma^2})$; n=number of observations; m= model size of beta
 (refer to *setupsimulationYisXbetaplussigmaZ.m)*

2. Use the routine iterativeST.m for iterative soft-thresholding and make a plot of the exact prediction error as a function of the number of iteration steps.

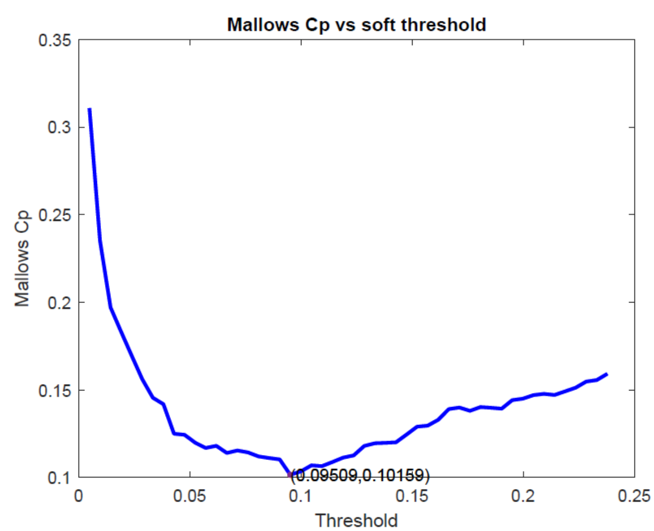


3. Construct a set of, say 50, thresholds between 0 and $\frac{\sqrt{2\ln n}\sigma}{5}$ (division by factor 5 after having observed that min PE error s way below uunivthr). Make a plot of the prediction error at the end of the iterative soft threshold procedure as a function of the threshold.



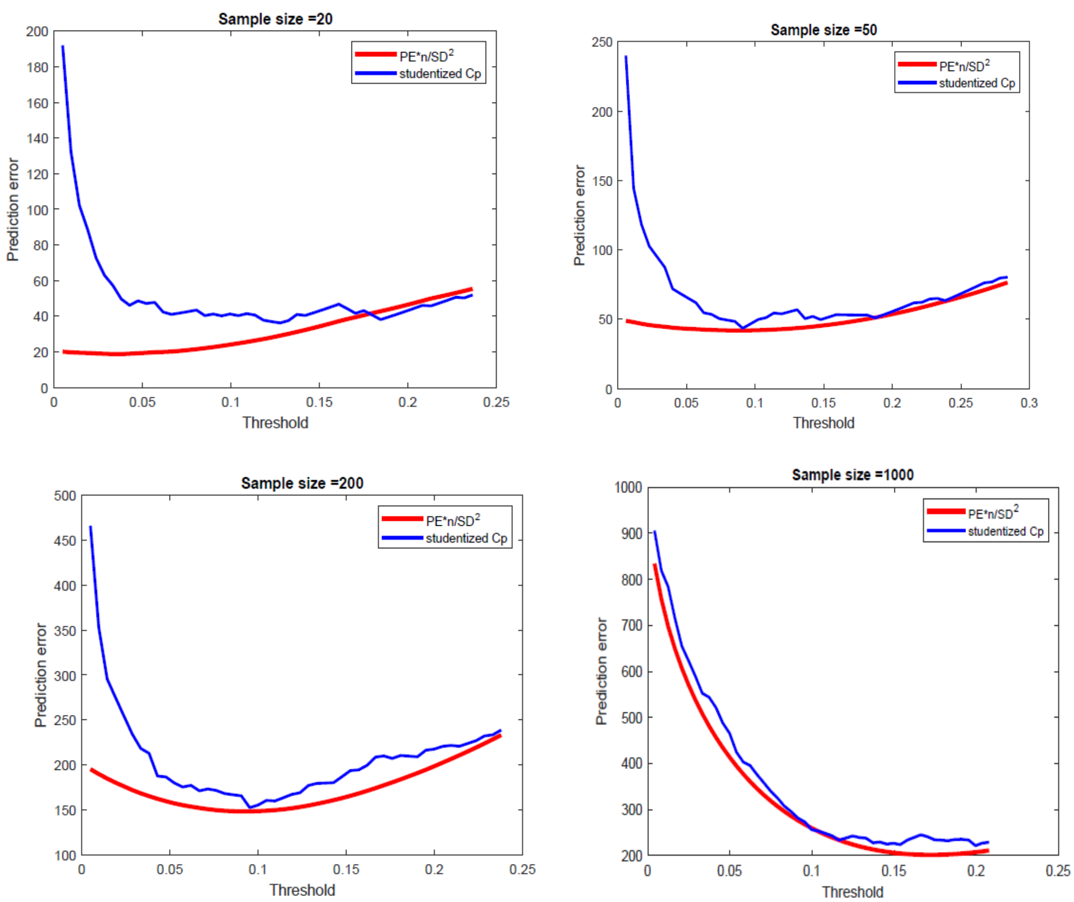
Also make a plot of the Mallows' Cp-value. Use the definition for Cp $\Delta(\widehat{\beta}_p) = \frac{1}{n}\mathrm{SS}_{\mathrm{E}p}(\widehat{\beta}_p) + \frac{2\nu_p}{n}\sigma^2 - \sigma^2$ and for the prediction error $\mathrm{PE}(\widehat{\beta}_p) = \frac{1}{n}E\left[\|\boldsymbol{X}\widehat{\boldsymbol{\beta}}_p - \boldsymbol{X}\boldsymbol{\beta}\|_2^2\right]$.

Note: assume that we do know the true model (β). So we compute the observed (not the expected) value of the error of the prediction

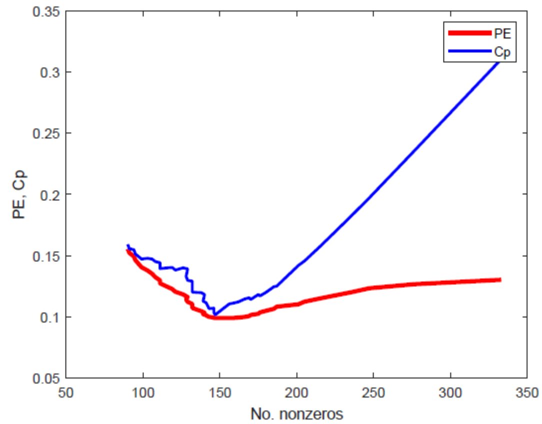


(a) Both above plots show the characteristic curve with a global minimum corresponding to the optimal trade-off between bias and variance (*by PE*), or between closeness and complexity (*by Cp*). The corresponding model sizes are 19 and 20 for using *soft threshold* and *Mallows' Cp*, respectively.

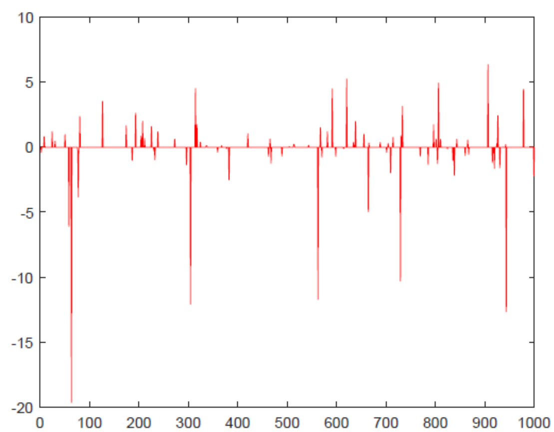
(b) Using the definitions above, we have $E\left(\Delta(\hat{\beta}_p)\right) = PE(\hat{\beta}_p)$. When using the studentized version of $Cp = \frac{SS_{Ep}}{\sigma^2} + 2p - n$, no such pointwise unbiasedness holds. Cp is pointwise consistent $Cp = \frac{n}{\sigma^2} PE(\hat{\beta}_p)$. Note: "Prediction error" in below plots are scaled errors.



4. Compute for every threshold, the number of nonzeros in the result, then plot Cp and prediction error versus number of nonzeros.



5. Use the routine LARS.m to estimate β. The estimated β values at one threshold is shown below,



And make a plot the Cp-value as a function of the number of nonzeros.