Data-driven Lambda for the LASSO

Shuofan Zhang

December 11, 2018

In this report, only the significant lags chosen by AR were included in the X matrix.

Since the plots for yr2, yr3, yr4 and yr5 are similar, so here I just plotted yr2.

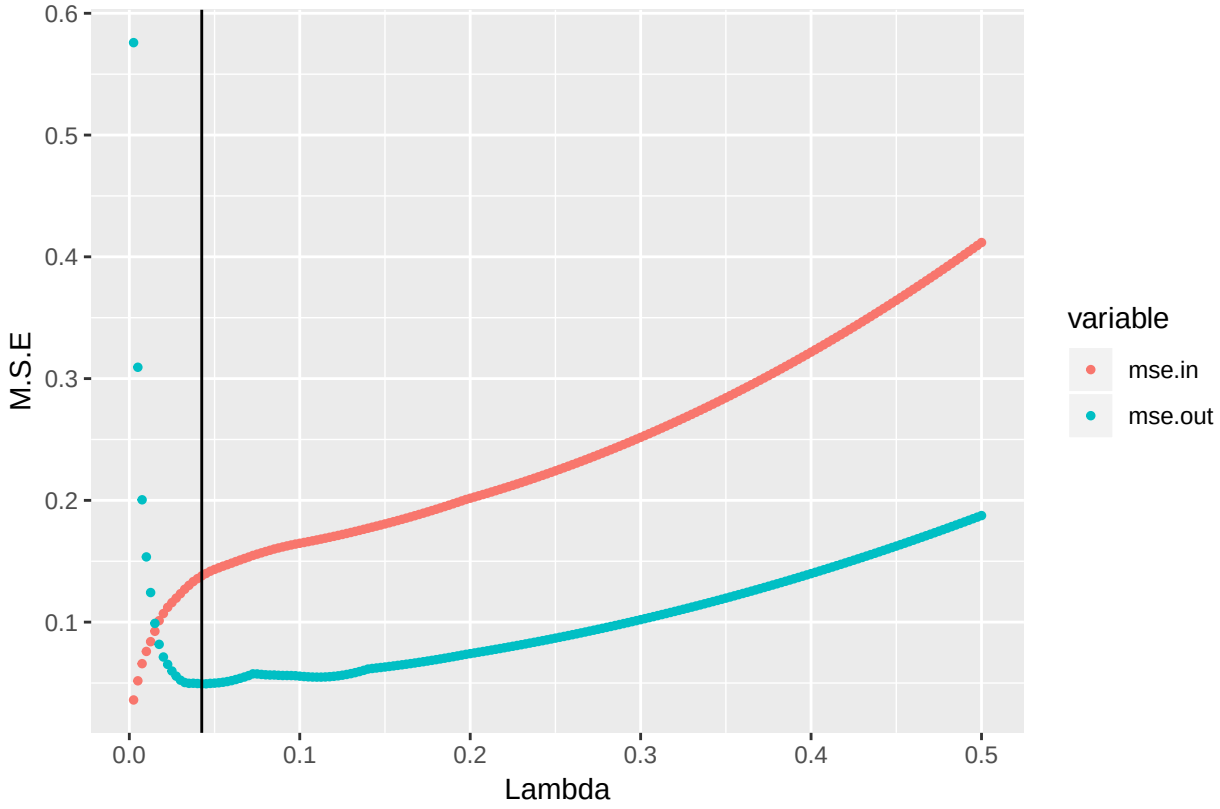
The M.S.E of the training dataset increases with λ, so it is not feasible to choose a λ according to the performance in training dataset.

Two solutions:

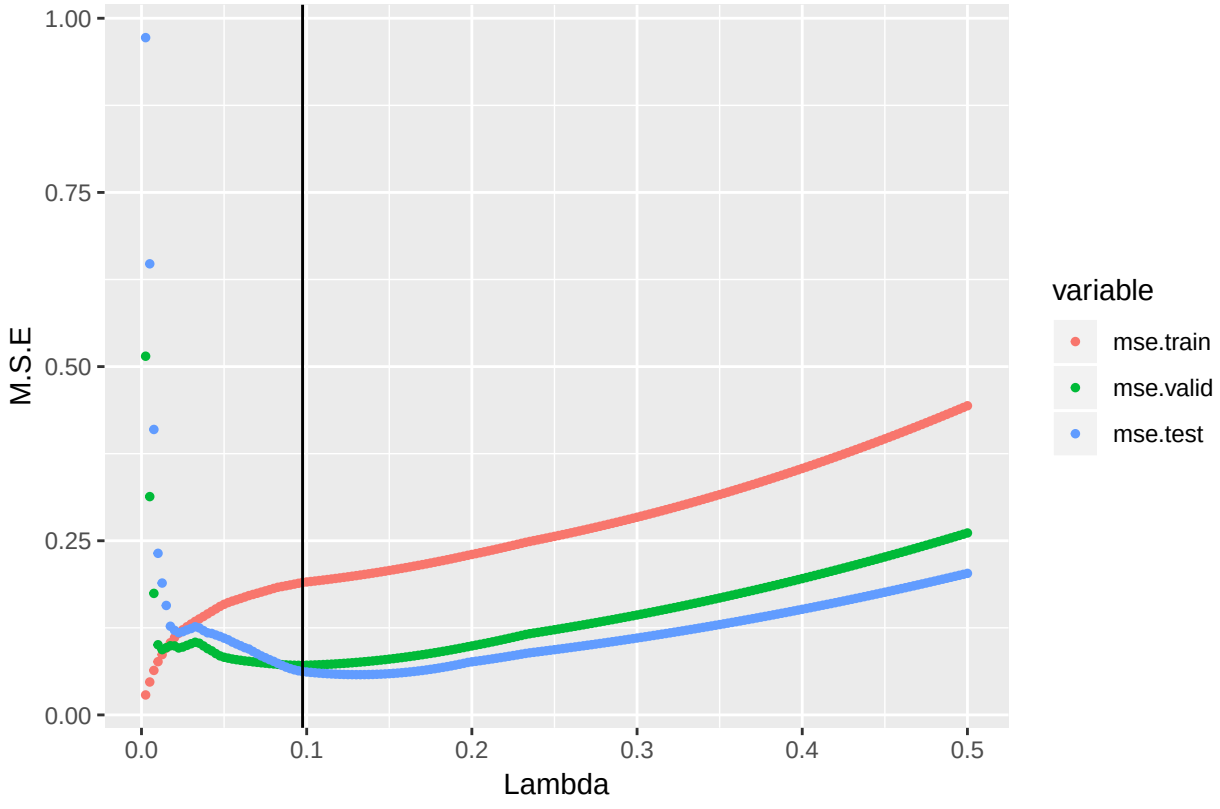
1, Same as before, separate the data set into training set (80%) and test set (20%); then choose λ according to the smallest M.S.E in the test set directly and use it for the comparison.

2, Separate the data set into three sets, training (60%), validation (20%) and test (20%); then choose λ according to the smallest M.S.E in the validation set, use the M.S.E from the test set for comparison.

- LASSO 1

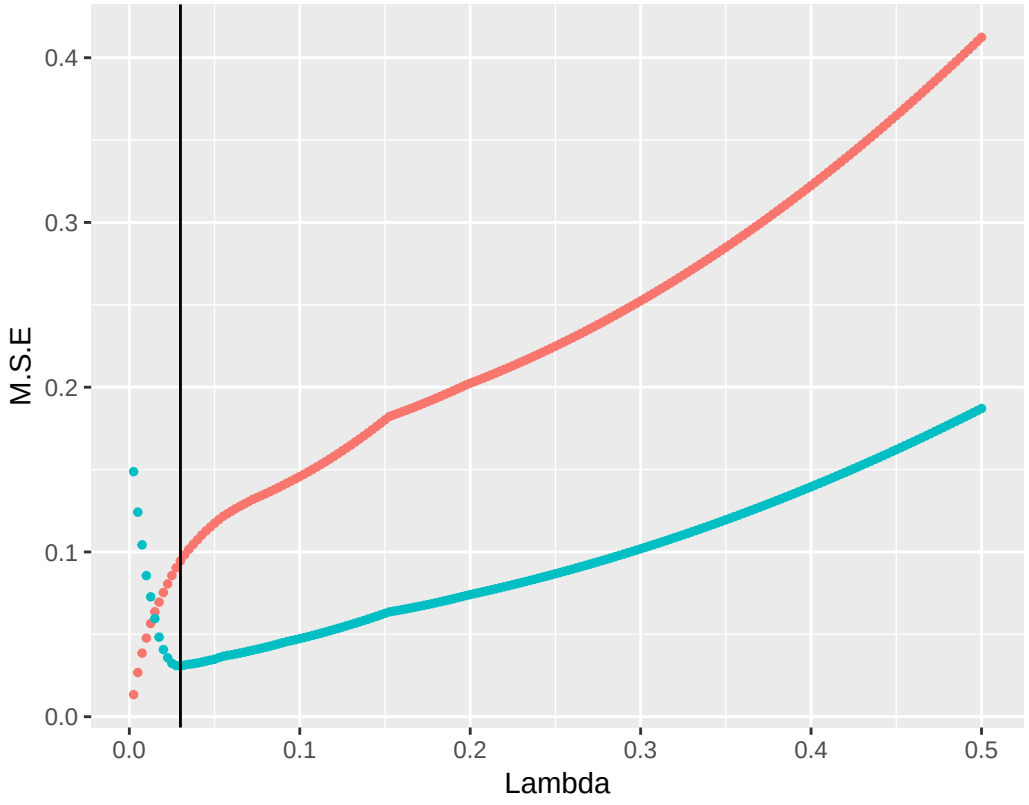


Path of M.S.E in two datasets against lambda, LASSO 1

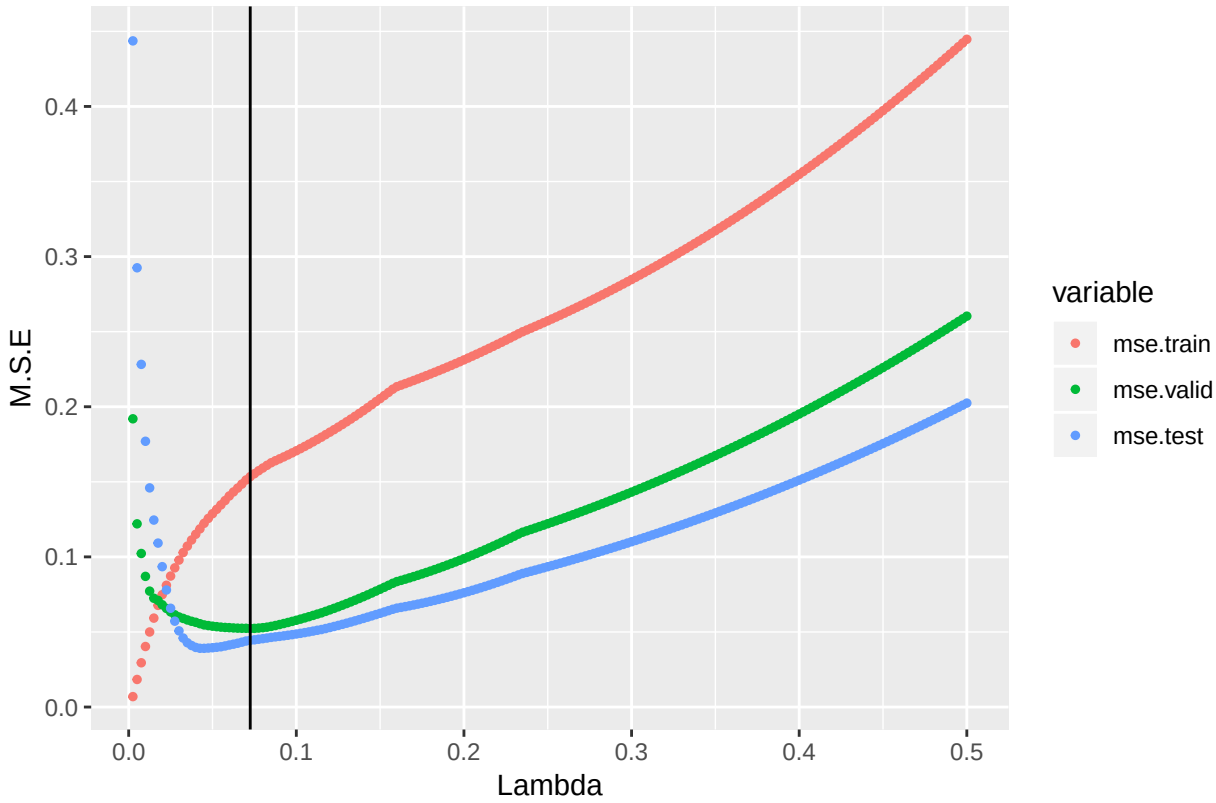


Path of M.S.E in three datasets against lambda, LASSO 1

- LASSO 2

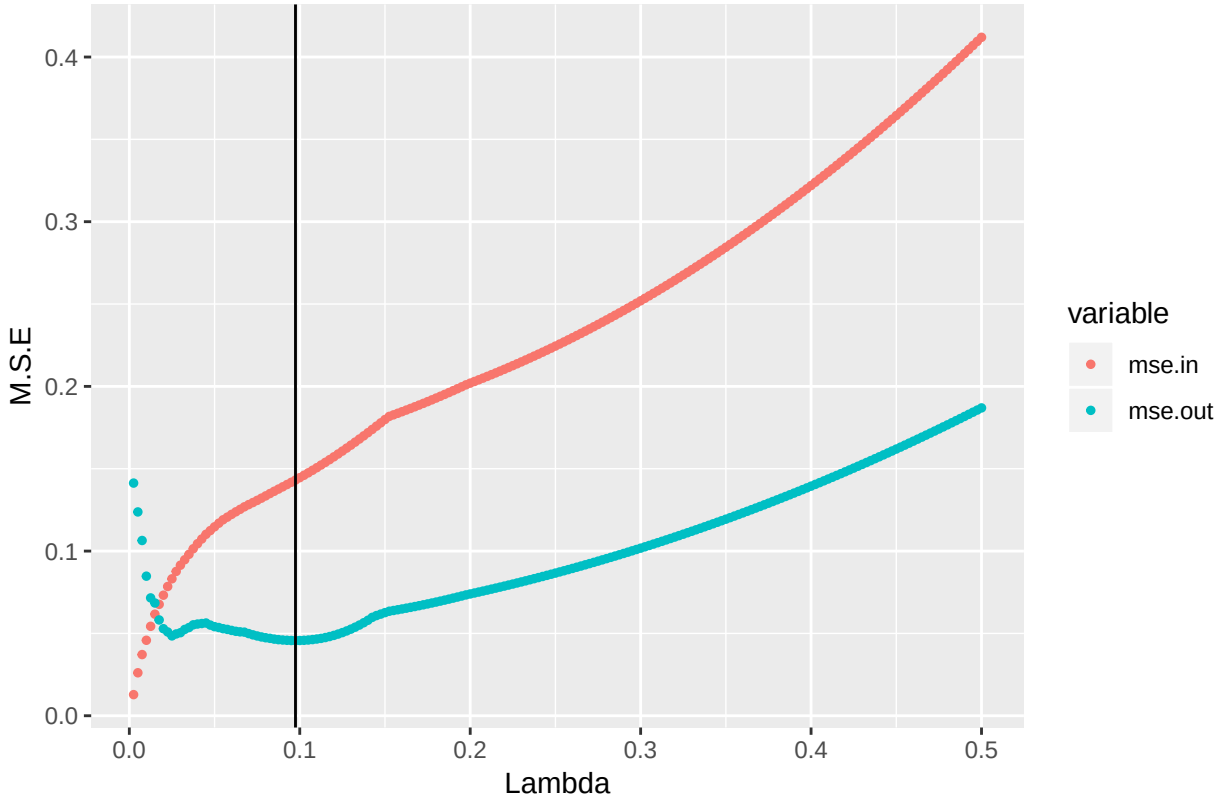


Path of M.S.E in two datasets against lambda, LASSO 2

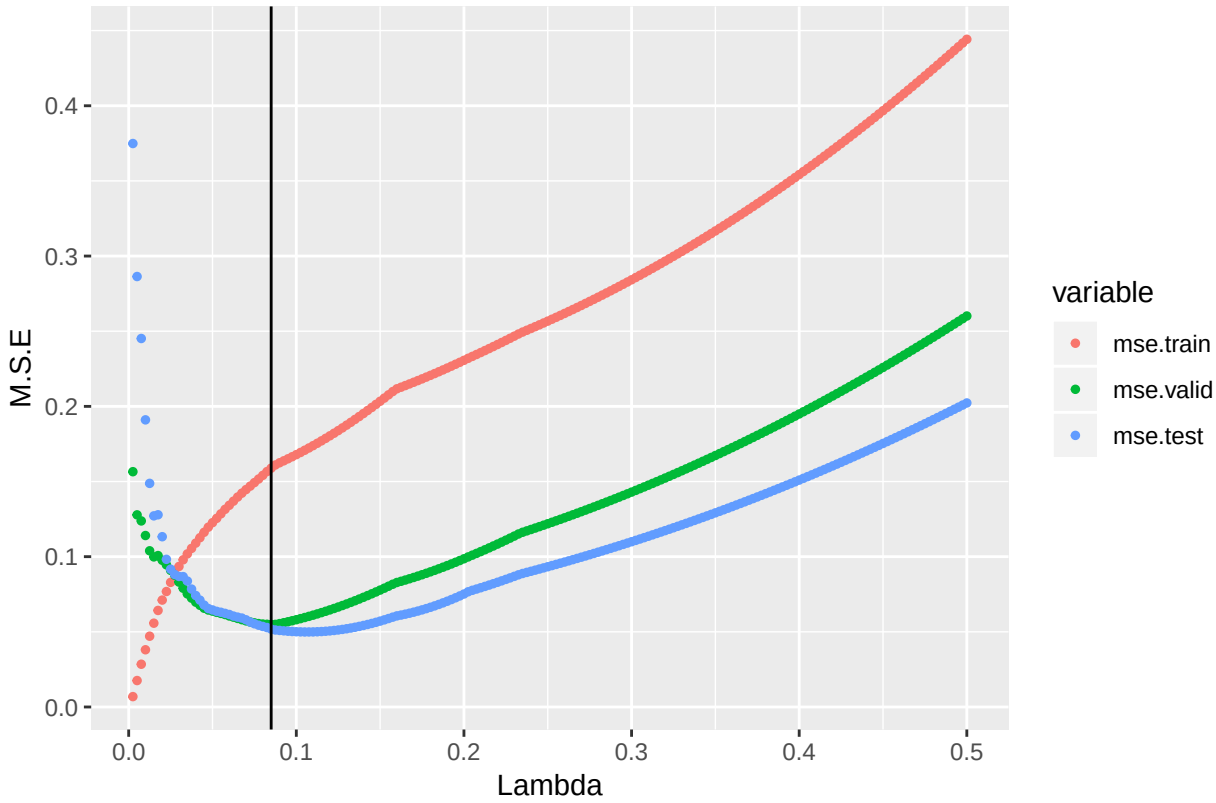


Path of M.S.E in three datasets against lambda, LASSO 2

- LASSO 3



Path of M.S.E in two datasets against lambda, LASSO 3



Path of M.S.E in three datasets against lambda, LASSO 3

Table 1: Variables selected by the LASSO with data driven lambda (valid) for yr2

variable	LASSO 1	LASSO 2	LASSO 3
ces015.lag1	-0.06302	NA	-0.081
pmcp.lag1	-0.08144	-0.07776	-0.06601
D.gmdcn.lag5	-3e-06	NA	NA
fclbmc.lag12	0.00039	NA	NA
yr.lag1	0.7669	0.8223	0.795
yr.lag5	NA	0.009655	NA
D.ypr.lag1	NA	-0.003463	-0.000973
D.a0m051.lag1	NA	-0.01647	-0.004059
pmcp.lag5	NA	-0.01831	-0.001411
D.fygt10.lag5	NA	-0.003416	-0.001061
D.fygt1.lag12	NA	0.001242	1.4e-05
D.fygt5.lag12	NA	0.04613	0.04731
D.fygt10.lag12	NA	0.04782	0.03405
D.fyff.lag13	NA	-0.004052	NA
exrus.lag1	NA	NA	0.002564
exruk.lag1	NA	NA	-0.002343