

Supplement to “Extended Comparisons of Best Subset Selection, Forward Stepwise Selection, and the Lasso”

Trevor Hastie

Robert Tibshirani

Ryan J. Tibshirani

This supplementary document contains plots from the simulation suite described in the paper “Extended Comparisons of Best Subset Selection, Forward Stepwise Selection, and the Lasso”. The plots in Section A precisely follow the simulation format described in the paper. Those in Section B follow an analogous format, except that the tuning has been done using an “oracle”, rather than a validation set as in Section A. Specifically, the tuning parameter for each method in each scenario is chosen to minimize the average risk over all of the repetitions. Sections A and B compare best subset selection, forward stepwise regression, the lasso and the relaxed lasso. In Sections C and D we have added three more methods to the comparisons: L0Learn-1, L0Learn-2 and SparseNet.

Contents

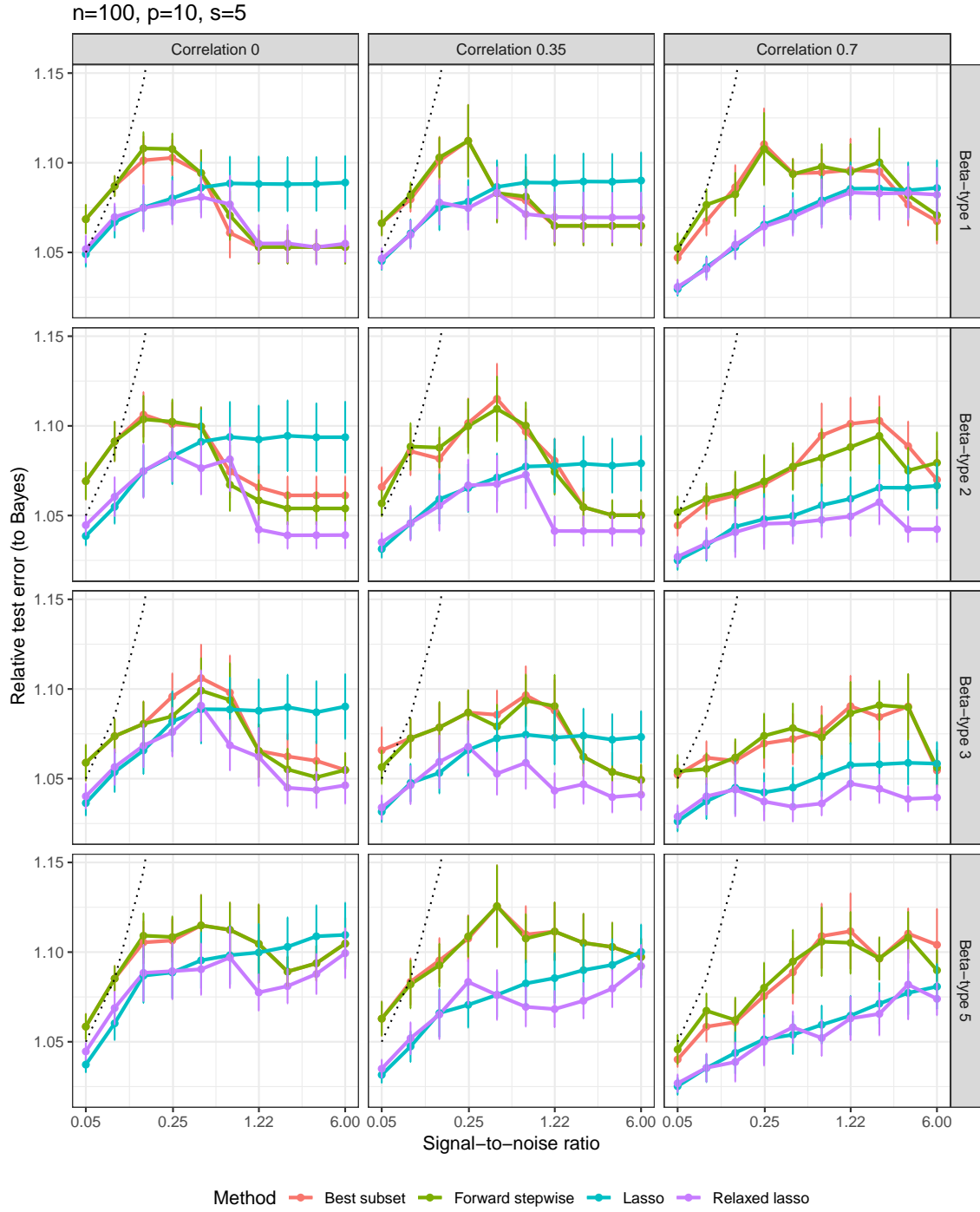
| | | |
|----------|--|-----------|
| A | Validation tuning | 4 |
| A.1 | Low setting: $n = 100, p = 10, s = 5$ | 4 |
| A.1.1 | Relative test error (to Bayes) | 4 |
| A.1.2 | Proportion of variance explained | 5 |
| A.1.3 | Number of nonzero coefficients | 6 |
| A.1.4 | F-score | 7 |
| A.2 | Medium setting: $n = 500, p = 100, s = 5$ | 8 |
| A.2.1 | Relative test error (to Bayes) | 8 |
| A.2.2 | Proportion of variance explained | 9 |
| A.2.3 | Number of nonzero coefficients | 10 |
| A.2.4 | F-score | 11 |
| A.3 | High-5 setting: $n = 50, p = 1000, s = 5$ | 12 |
| A.3.1 | Relative test error (to Bayes) | 12 |
| A.3.2 | Proportion of variance explained | 13 |
| A.3.3 | Number of nonzero coefficients | 14 |
| A.3.4 | F-score | 15 |
| A.4 | High-10 setting: $n = 100, p = 1000, s = 10$ | 16 |
| A.4.1 | Relative test error (to Bayes) | 16 |
| A.4.2 | Proportion of variance explained | 17 |
| A.4.3 | Number of nonzero coefficients | 18 |
| A.4.4 | F-score | 19 |
| B | Oracle tuning | 20 |
| B.1 | Low setting: $n = 100, p = 10, s = 5$ | 20 |
| B.1.1 | Relative test error (to Bayes) | 20 |
| B.1.2 | Proportion of variance explained | 21 |
| B.1.3 | Number of nonzero coefficients | 22 |
| B.1.4 | F-score | 23 |
| B.2 | Medium setting: $n = 500, p = 100, s = 5$ | 24 |
| B.2.1 | Relative test error (to Bayes) | 24 |
| B.2.2 | Proportion of variance explained | 25 |
| B.2.3 | Number of nonzero coefficients | 26 |
| B.2.4 | F-score | 27 |
| B.3 | High-5 setting: $n = 50, p = 1000, s = 5$ | 28 |
| B.3.1 | Relative test error (to Bayes) | 28 |
| B.3.2 | Proportion of variance explained | 29 |
| B.3.3 | Number of nonzero coefficients | 30 |
| B.3.4 | F-score | 31 |
| B.4 | High-10 setting: $n = 100, p = 1000, s = 10$ | 32 |
| B.4.1 | Relative test error (to Bayes) | 32 |
| B.4.2 | Proportion of variance explained | 33 |
| B.4.3 | Number of nonzero coefficients | 34 |
| B.4.4 | F-score | 35 |
| C | Validation tuning: now including L0Learn-1, L0Learn-2 and SparseNet | 36 |
| C.1 | Low setting: $n = 100, p = 10, s = 5$ | 36 |
| C.1.1 | Relative test error (to Bayes) | 36 |
| C.1.2 | Proportion of variance explained | 37 |
| C.1.3 | Number of nonzero coefficients | 38 |

| | | |
|----------|--|-----------|
| C.1.4 | F-score | 39 |
| C.2 | Medium setting: $n = 500, p = 100, s = 5$ | 40 |
| C.2.1 | Relative test error (to Bayes) | 40 |
| C.2.2 | Proportion of variance explained | 41 |
| C.2.3 | Number of nonzero coefficients | 42 |
| C.2.4 | F-score | 43 |
| C.3 | High-5 setting: $n = 50, p = 1000, s = 5$ | 44 |
| C.3.1 | Relative test error (to Bayes) | 44 |
| C.3.2 | Proportion of variance explained | 45 |
| C.3.3 | Number of nonzero coefficients | 46 |
| C.3.4 | F-score | 47 |
| C.4 | High-10 setting: $n = 100, p = 1000, s = 10$ | 48 |
| C.4.1 | Relative test error (to Bayes) | 48 |
| C.4.2 | Proportion of variance explained | 49 |
| C.4.3 | Number of nonzero coefficients | 50 |
| C.4.4 | F-score | 51 |
| D | Oracle tuning: now including L0Learn-1, L0Learn-2 and SparseNet | 52 |
| D.1 | Low setting: $n = 100, p = 10, s = 5$ | 52 |
| D.1.1 | Relative test error (to Bayes) | 52 |
| D.1.2 | Proportion of variance explained | 53 |
| D.1.3 | Number of nonzero coefficients | 54 |
| D.1.4 | F-score | 55 |
| D.2 | Medium setting: $n = 500, p = 100, s = 5$ | 56 |
| D.2.1 | Relative test error (to Bayes) | 56 |
| D.2.2 | Proportion of variance explained | 57 |
| D.2.3 | Number of nonzero coefficients | 58 |
| D.2.4 | F-score | 59 |
| D.3 | High-5 setting: $n = 50, p = 1000, s = 5$ | 60 |
| D.3.1 | Relative test error (to Bayes) | 60 |
| D.3.2 | Proportion of variance explained | 61 |
| D.3.3 | Number of nonzero coefficients | 62 |
| D.3.4 | F-score | 63 |
| D.4 | High-10 setting: $n = 100, p = 1000, s = 10$ | 64 |
| D.4.1 | Relative risk (to null model) | 64 |
| D.4.2 | Relative test error (to Bayes) | 64 |
| D.4.3 | Proportion of variance explained | 65 |
| D.4.4 | Number of nonzero coefficients | 66 |
| D.4.5 | F-score | 67 |

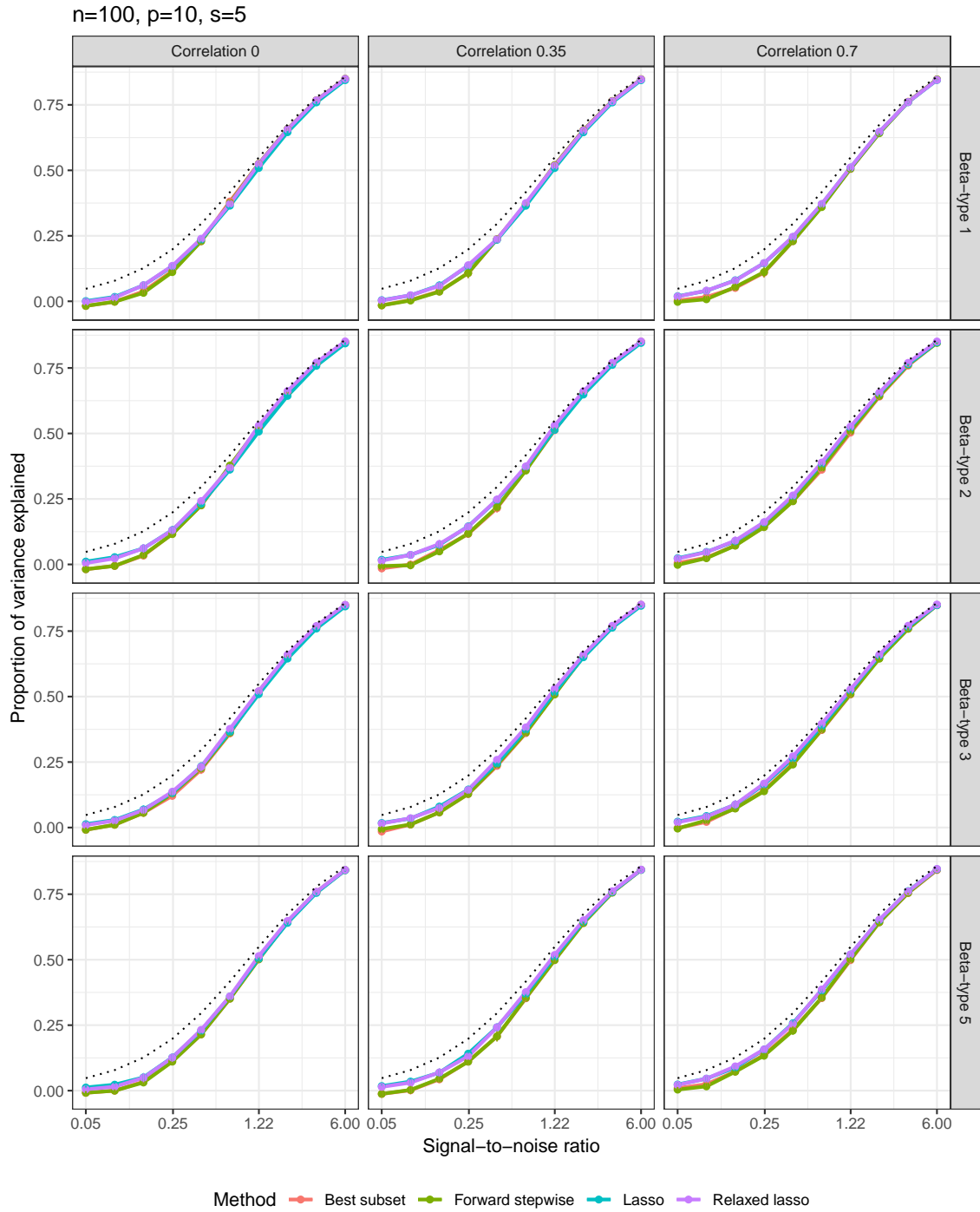
A Validation tuning

A.1 Low setting: $n = 100$, $p = 10$, $s = 5$

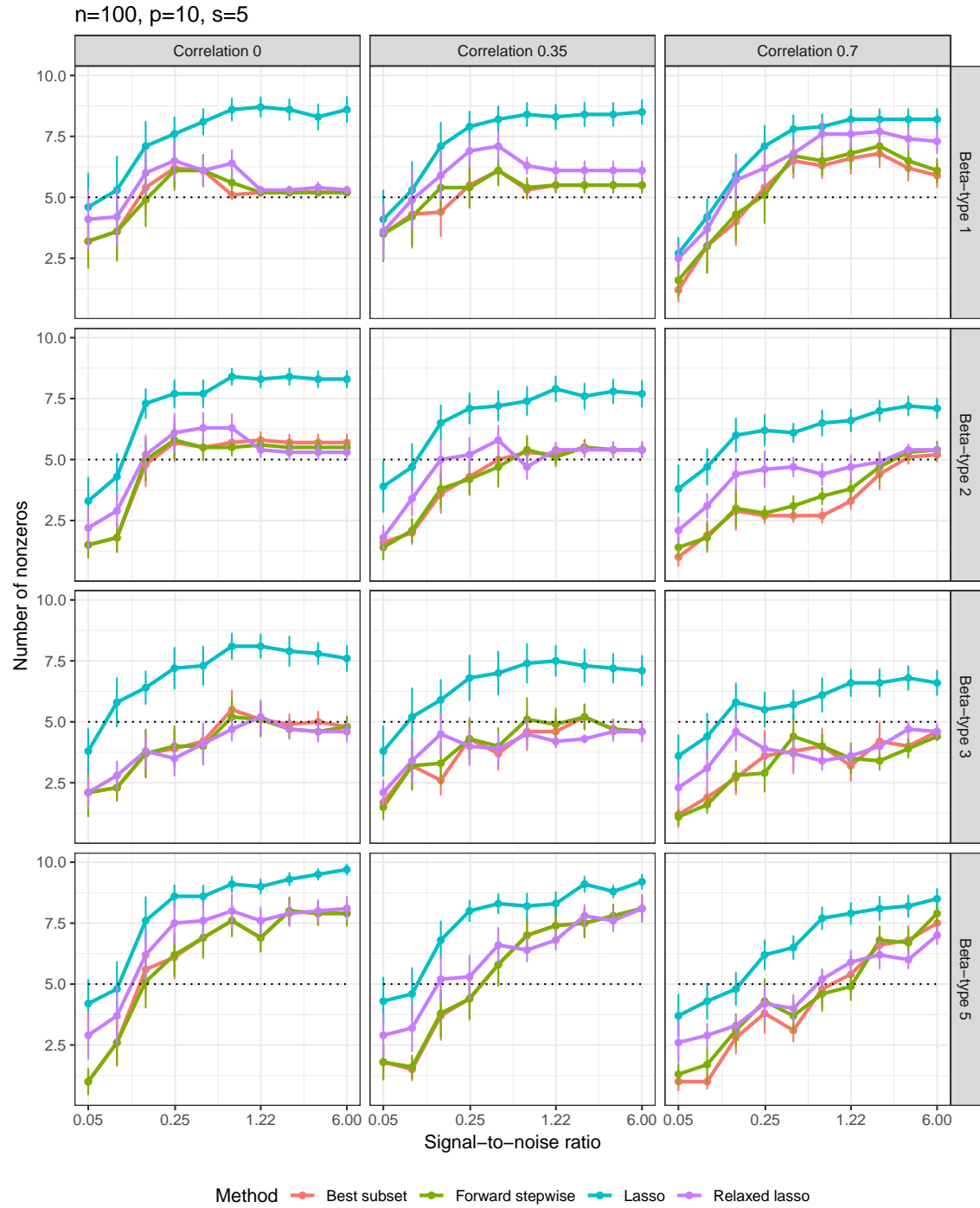
A.1.1 Relative test error (to Bayes)



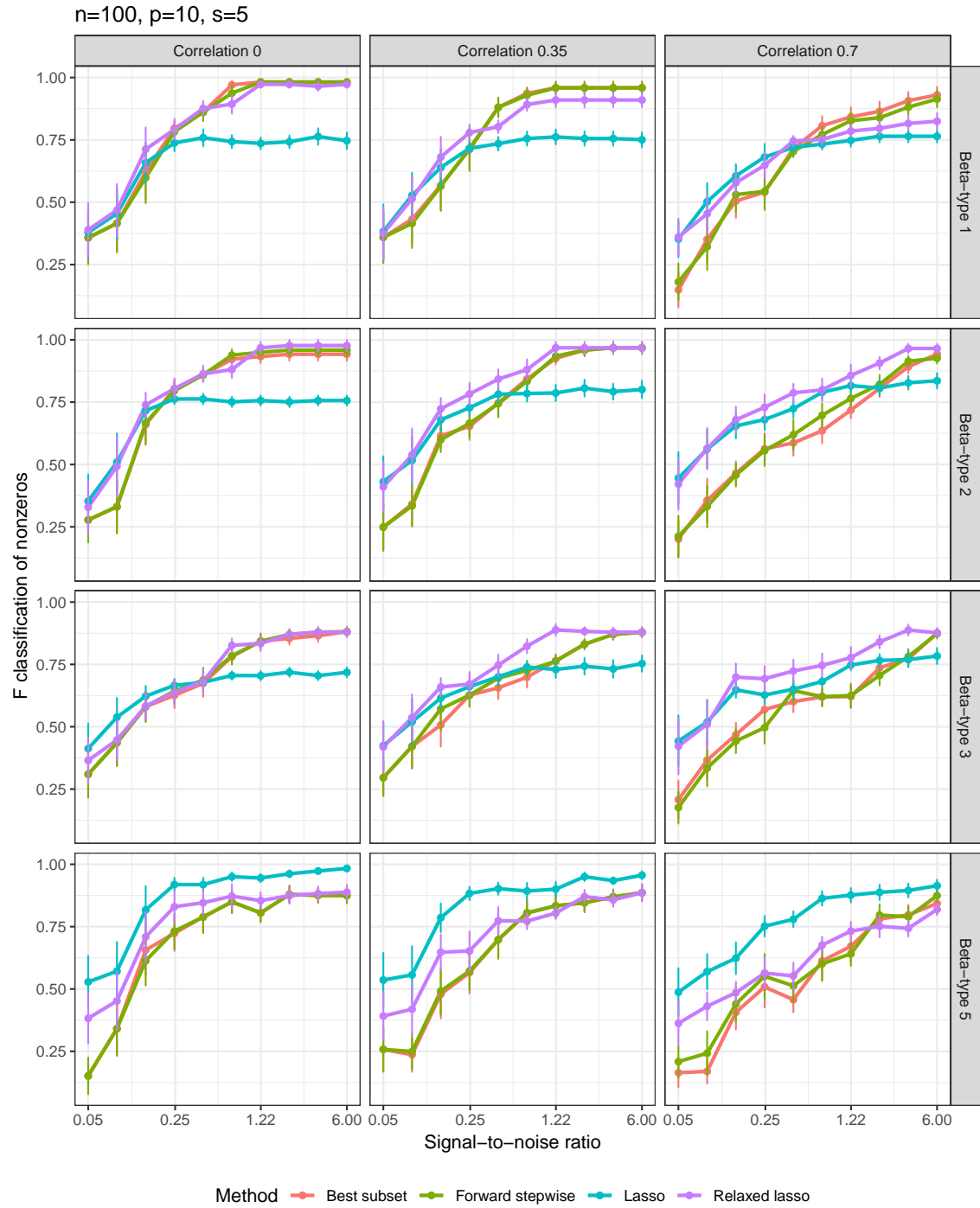
A.1.2 Proportion of variance explained



A.1.3 Number of nonzero coefficients

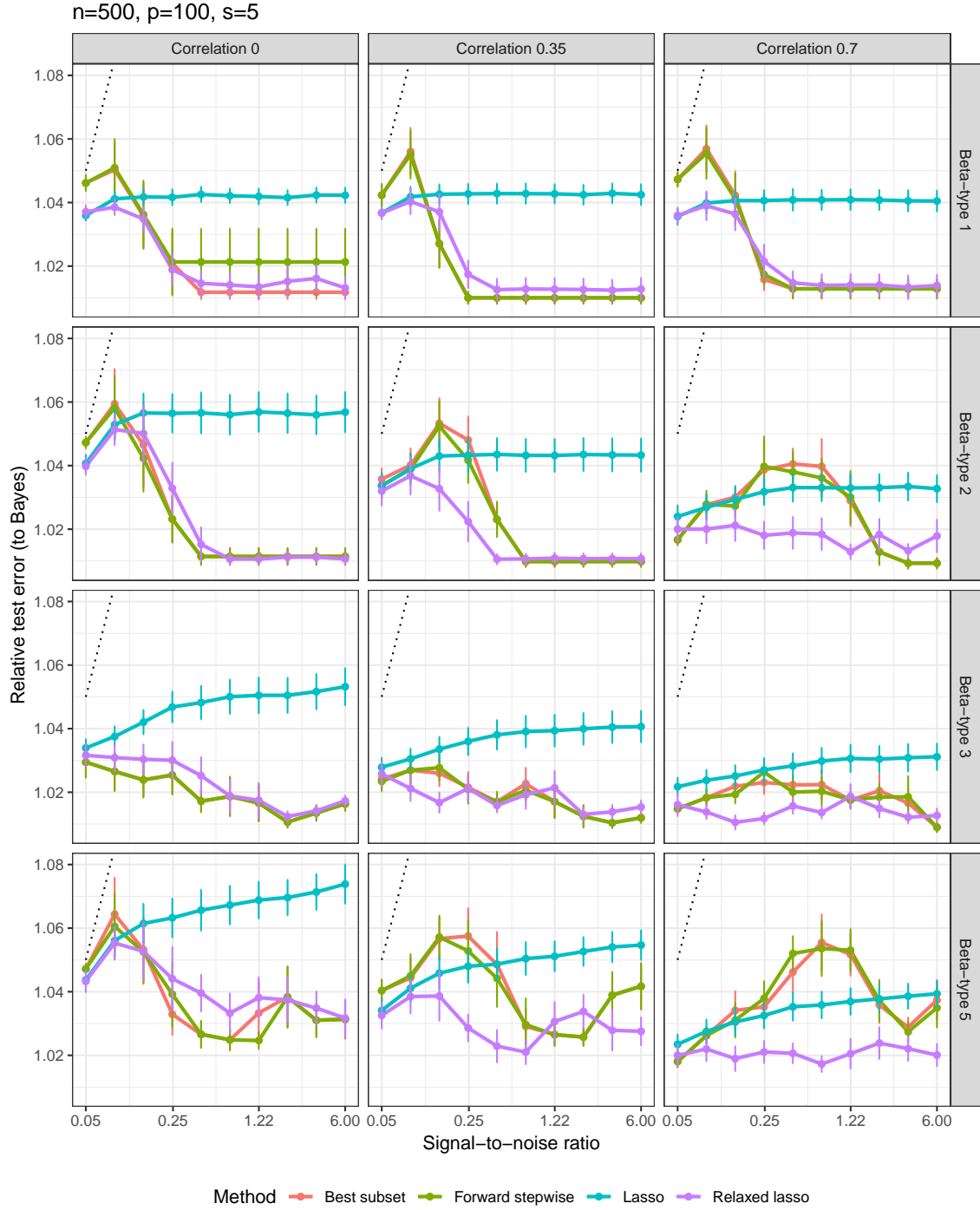


A.1.4 F-score

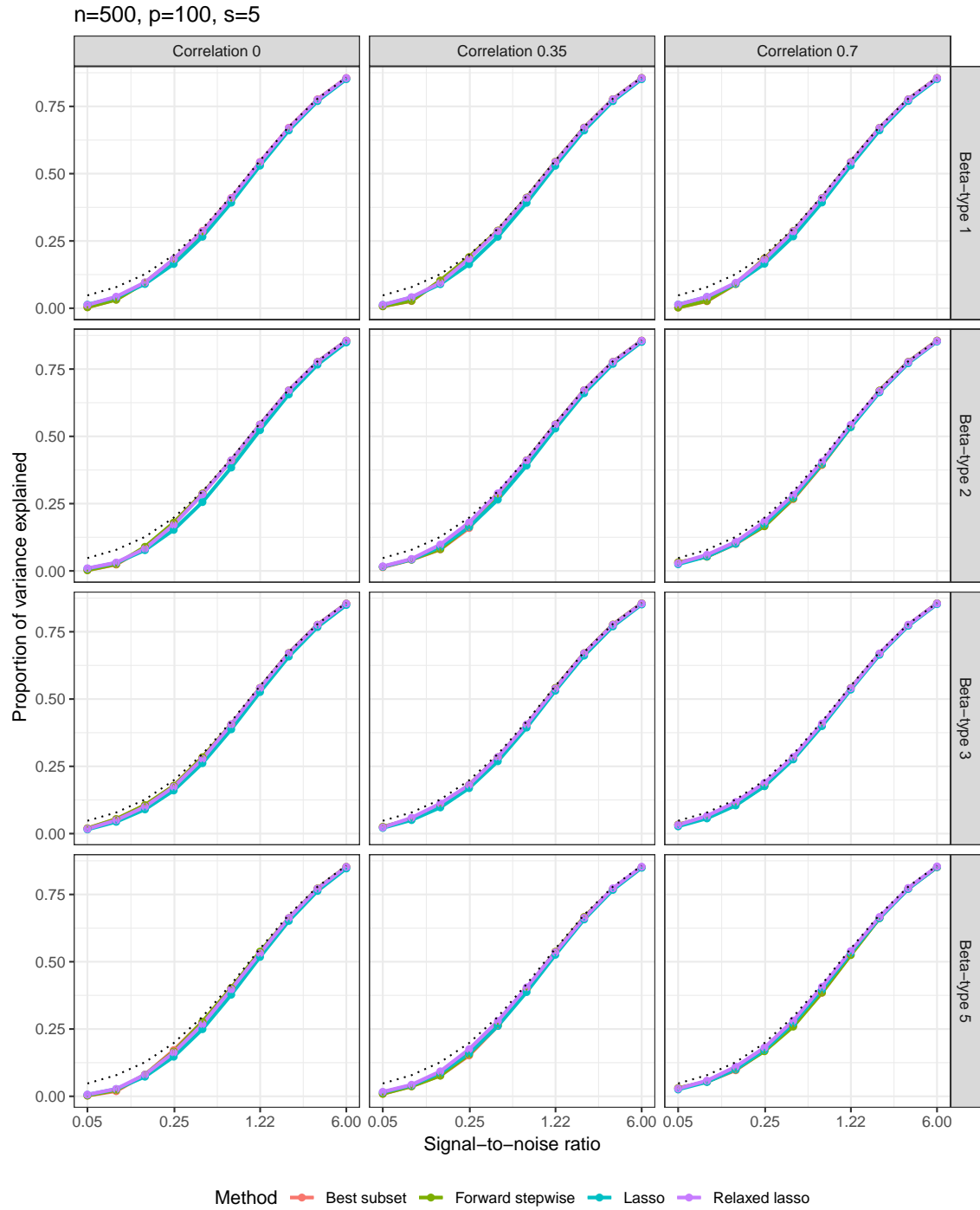


A.2 Medium setting: $n = 500$, $p = 100$, $s = 5$

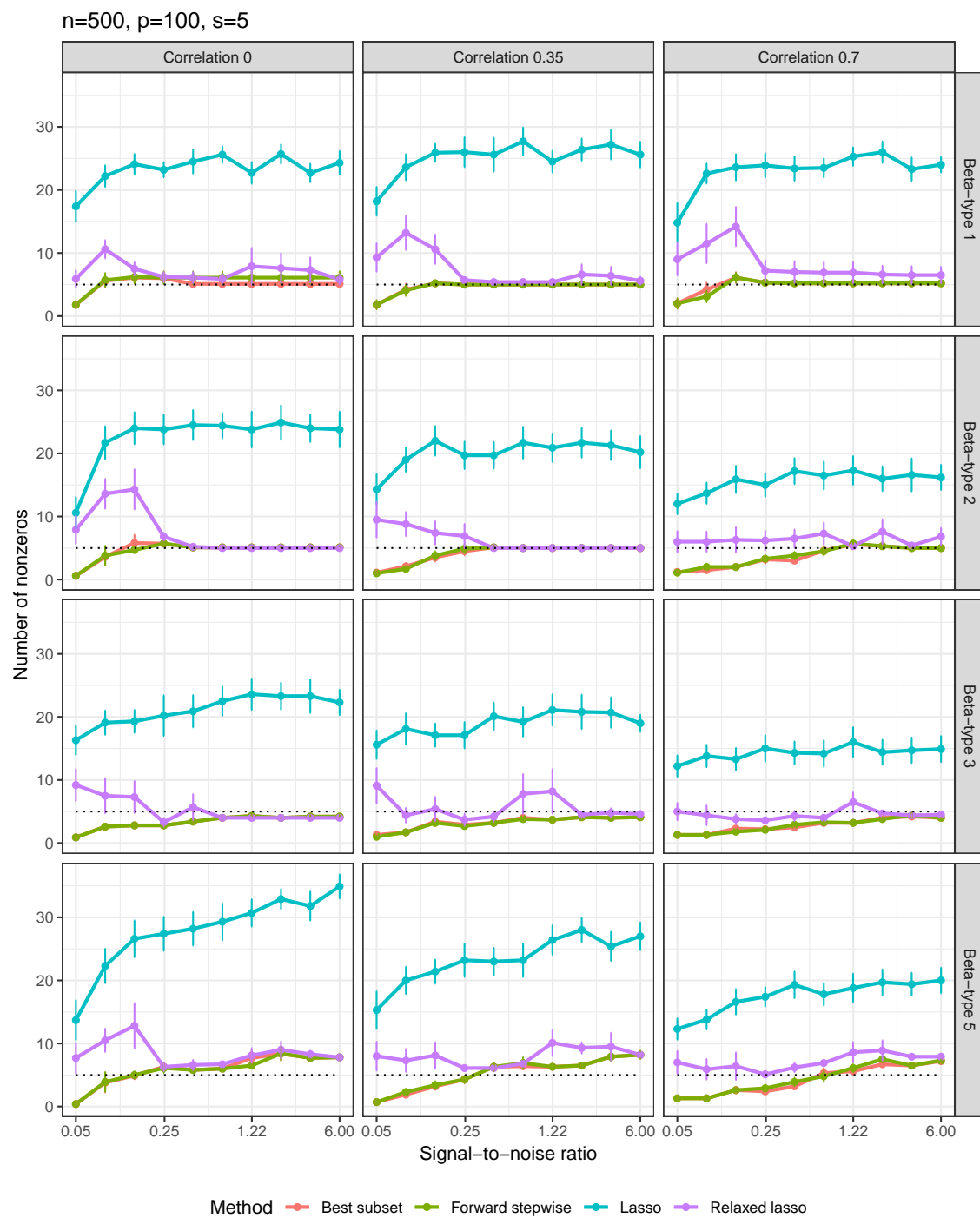
A.2.1 Relative test error (to Bayes)



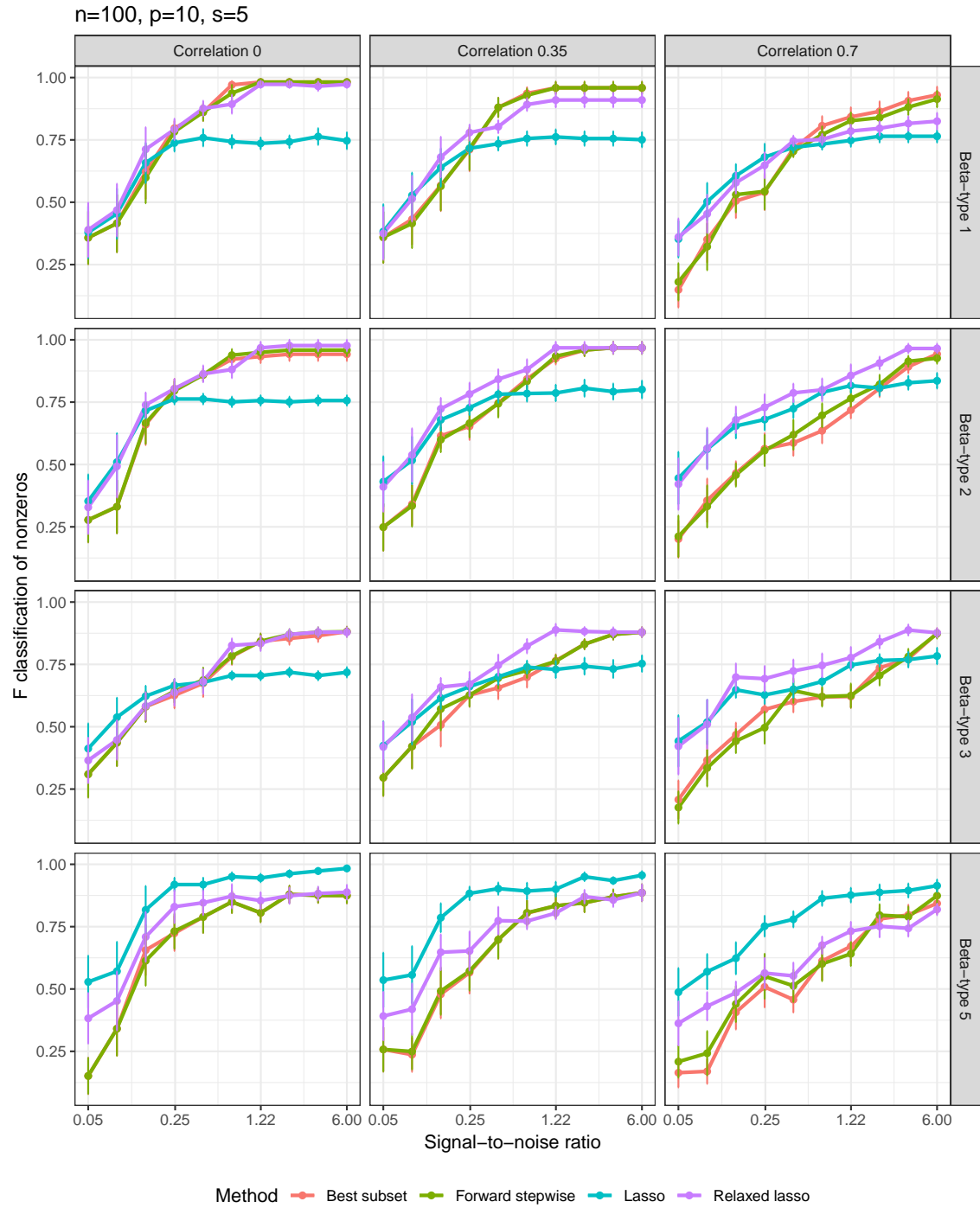
A.2.2 Proportion of variance explained



A.2.3 Number of nonzero coefficients

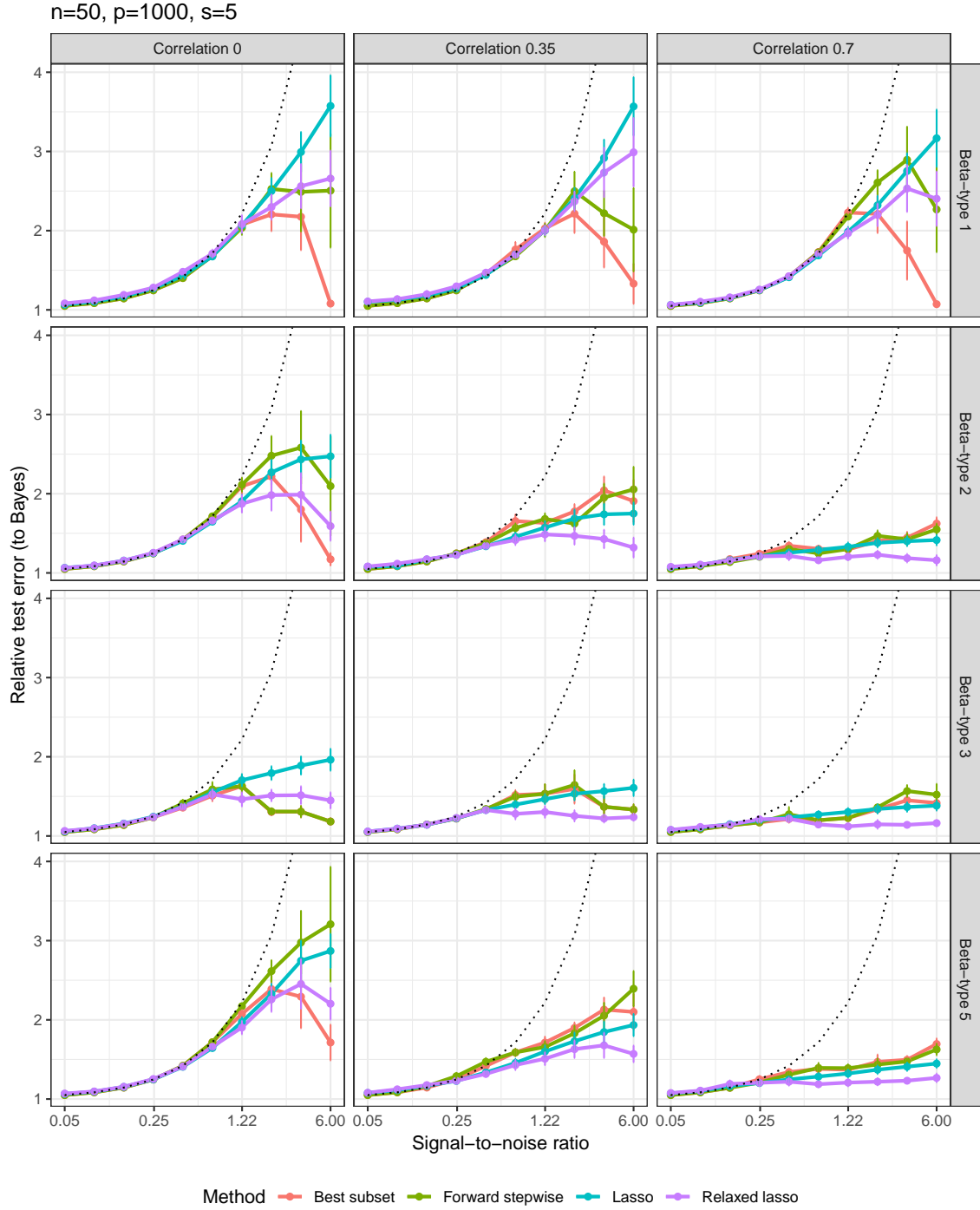


A.2.4 F-score

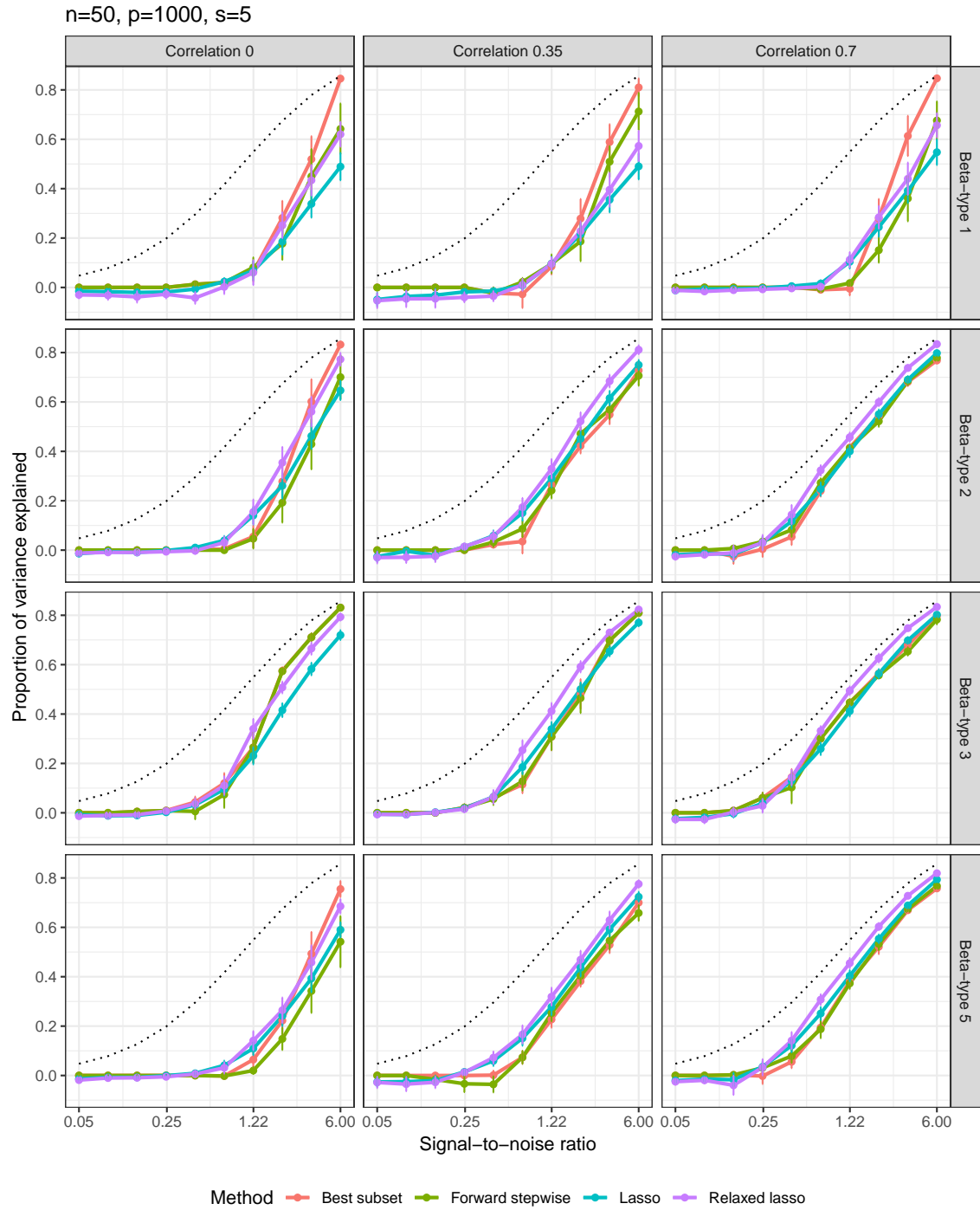


A.3 High-5 setting: $n = 50$, $p = 1000$, $s = 5$

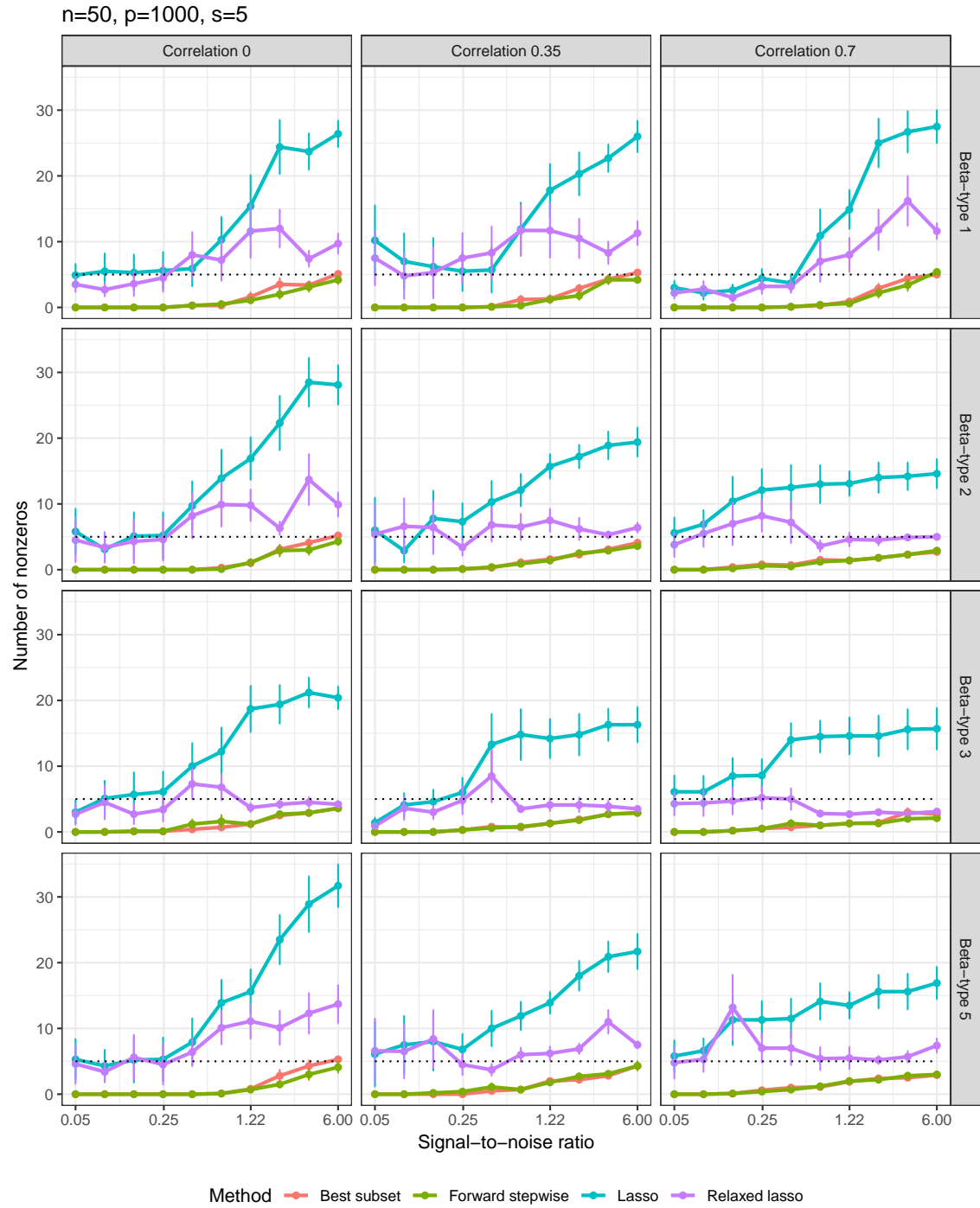
A.3.1 Relative test error (to Bayes)



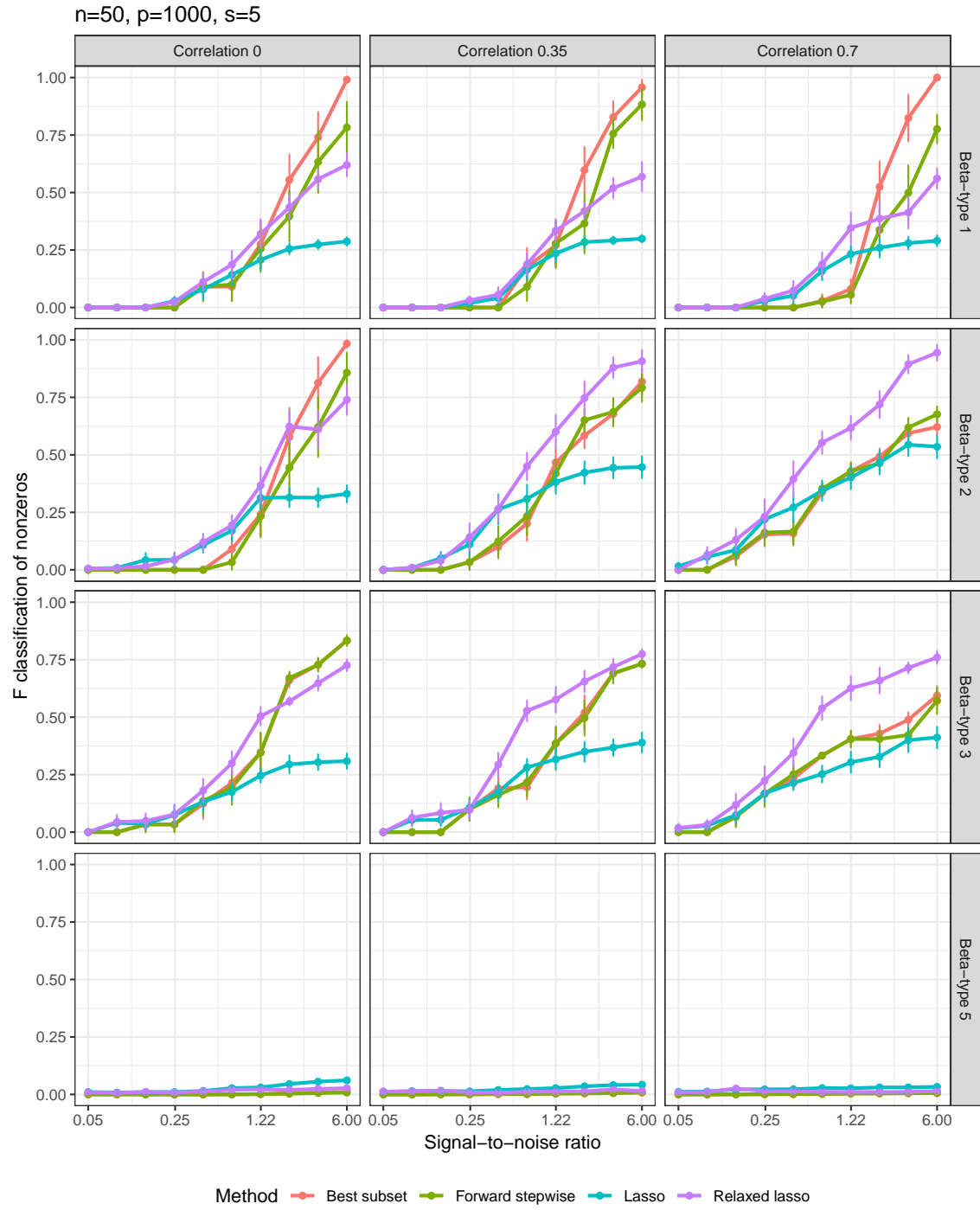
A.3.2 Proportion of variance explained



A.3.3 Number of nonzero coefficients

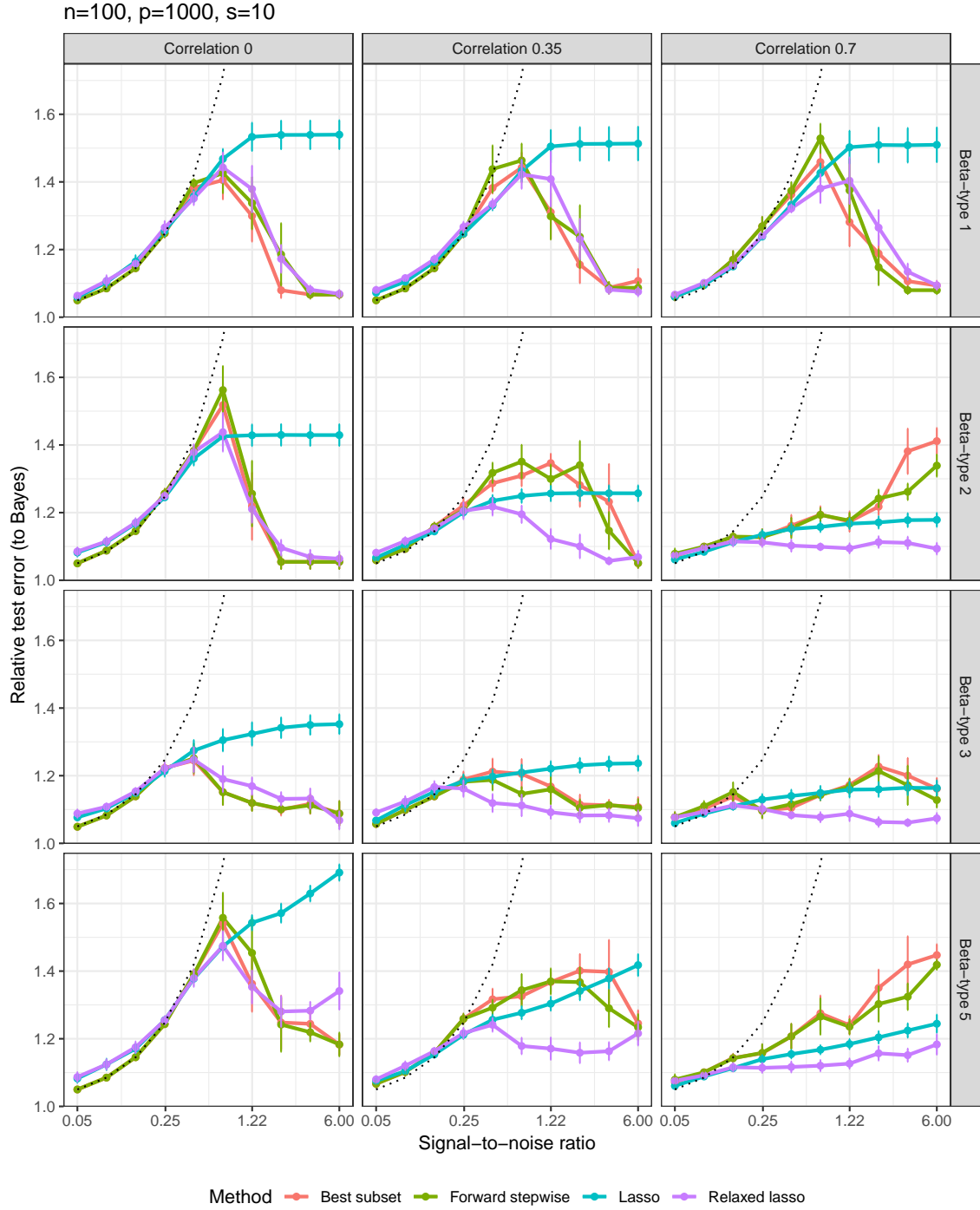


A.3.4 F-score

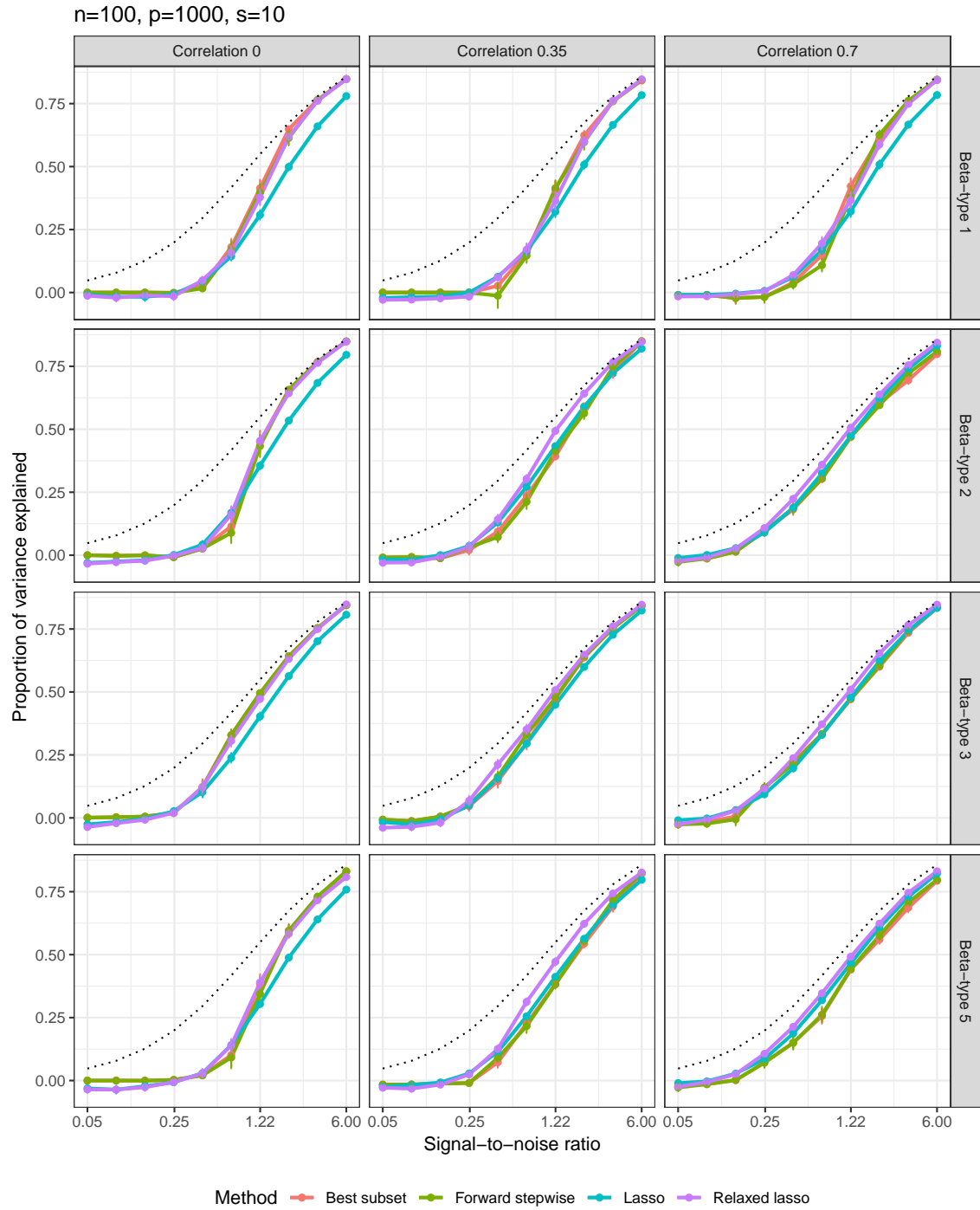


A.4 High-10 setting: $n = 100, p = 1000, s = 10$

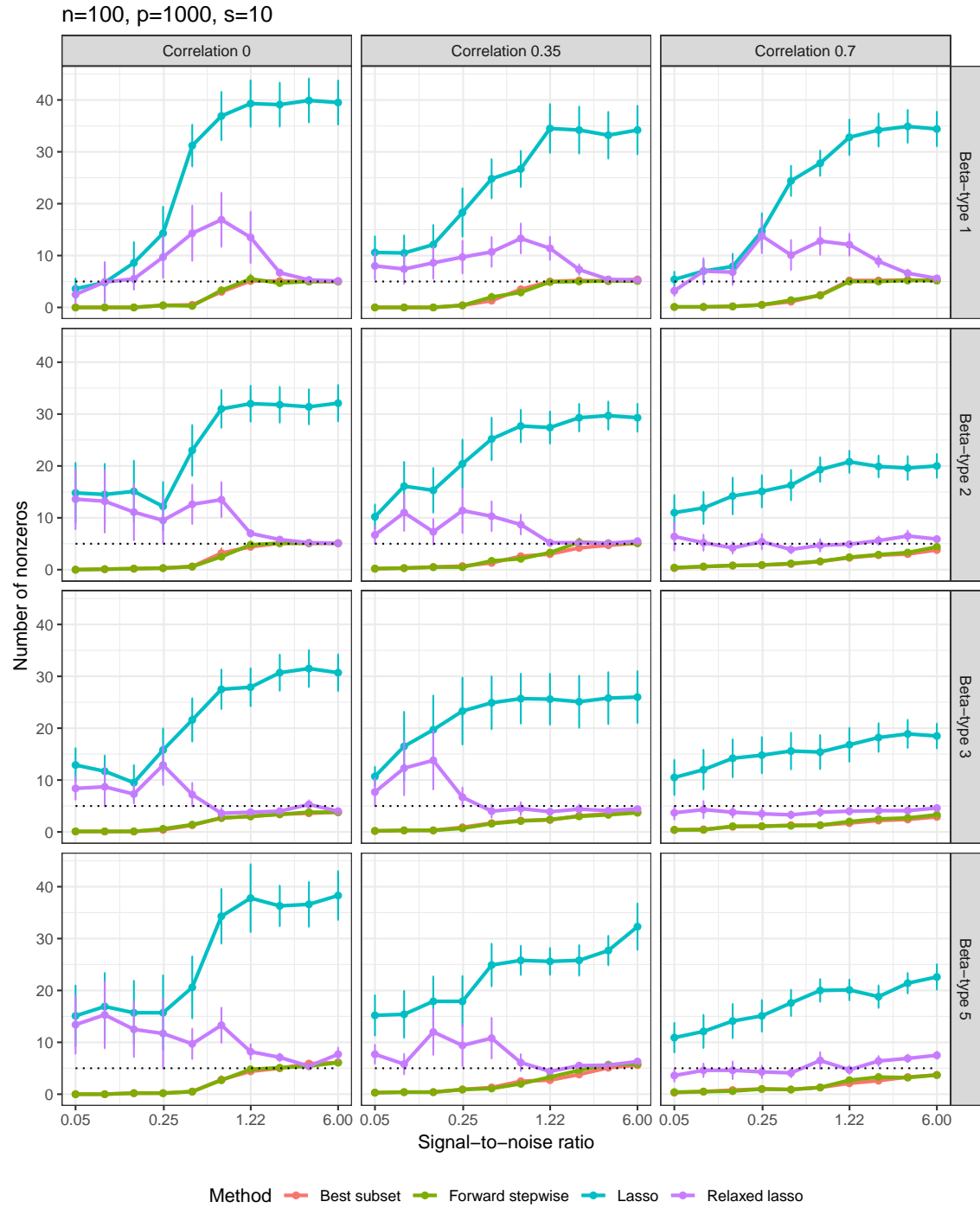
A.4.1 Relative test error (to Bayes)



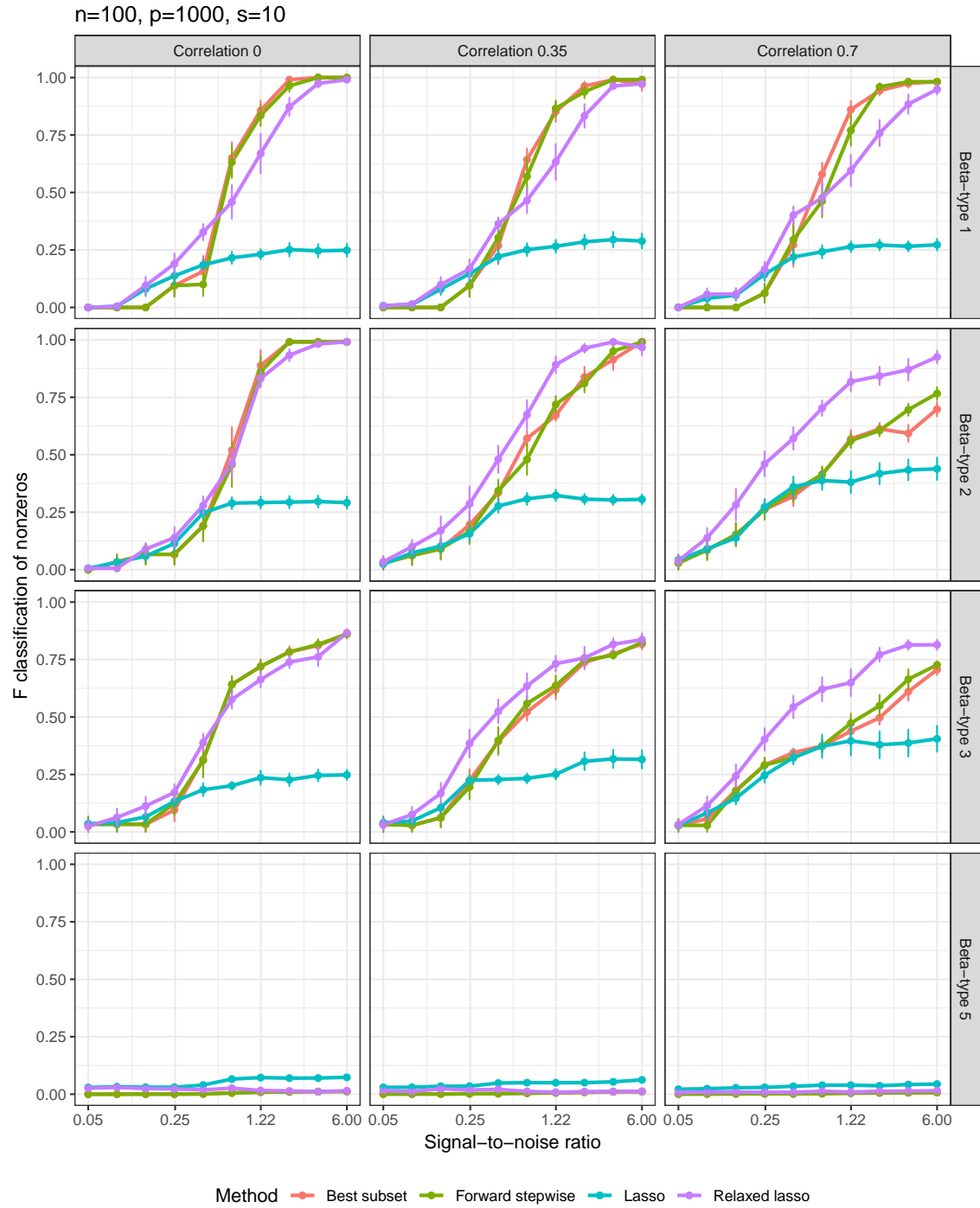
A.4.2 Proportion of variance explained



A.4.3 Number of nonzero coefficients



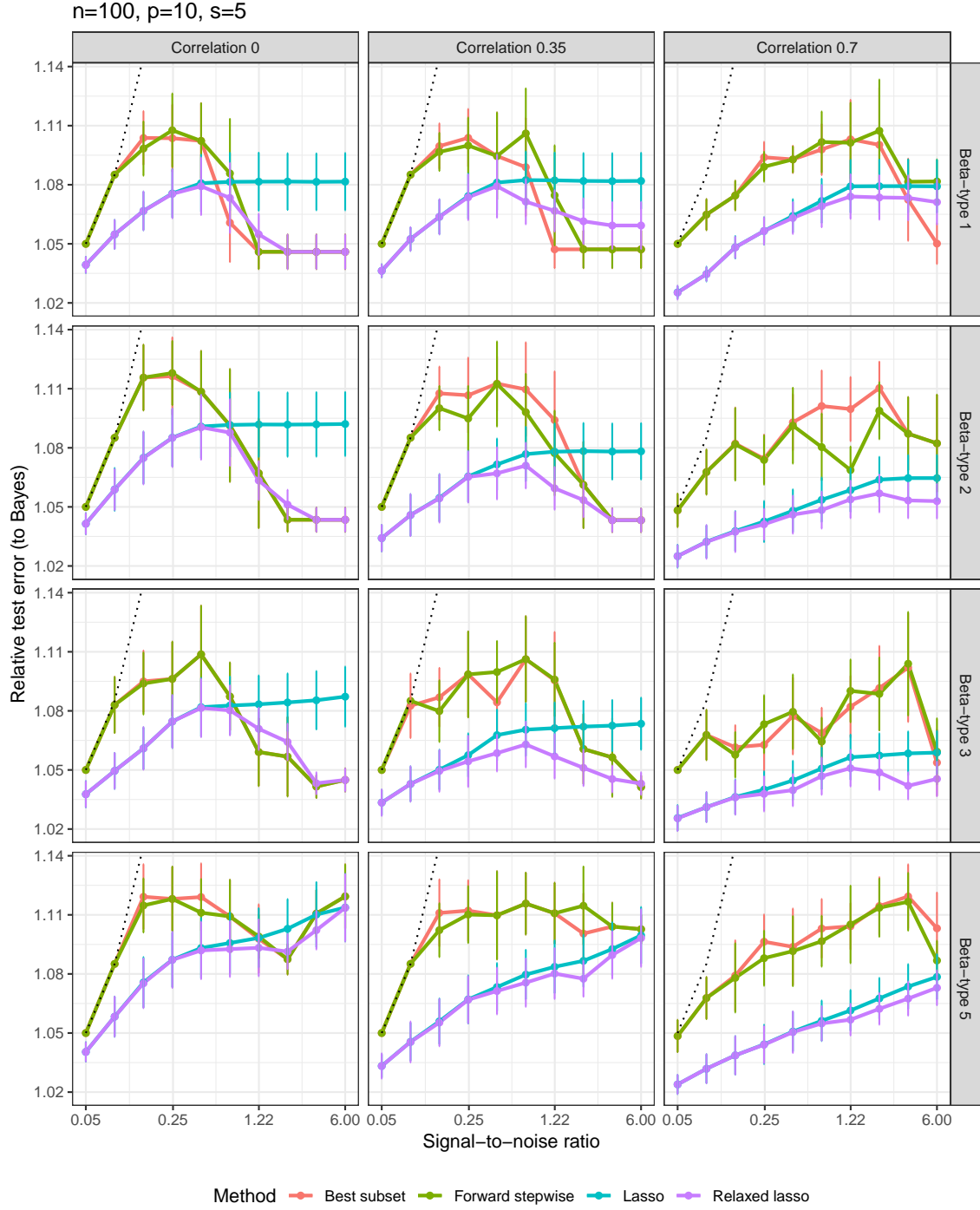
A.4.4 F-score



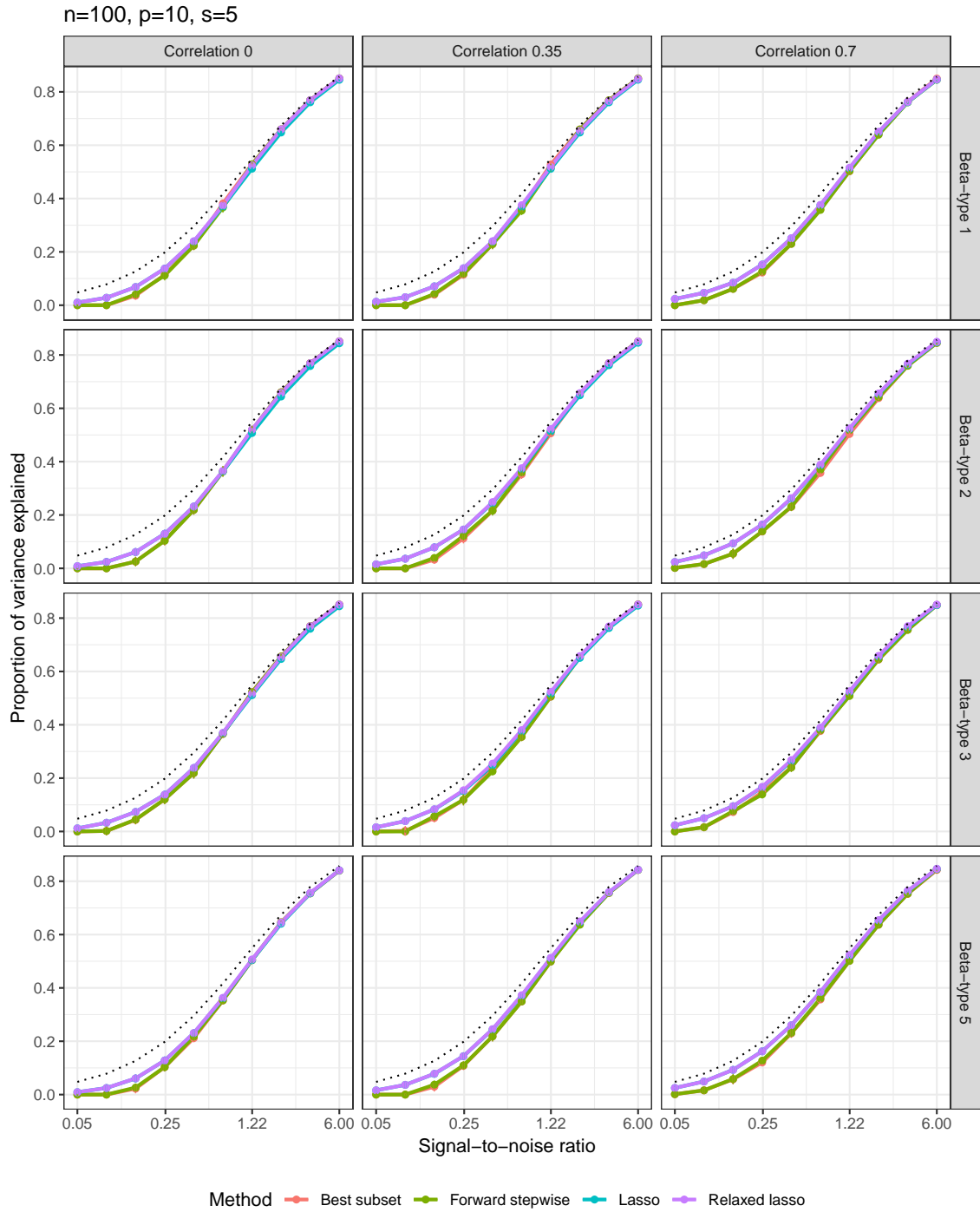
B Oracle tuning

B.1 Low setting: $n = 100, p = 10, s = 5$

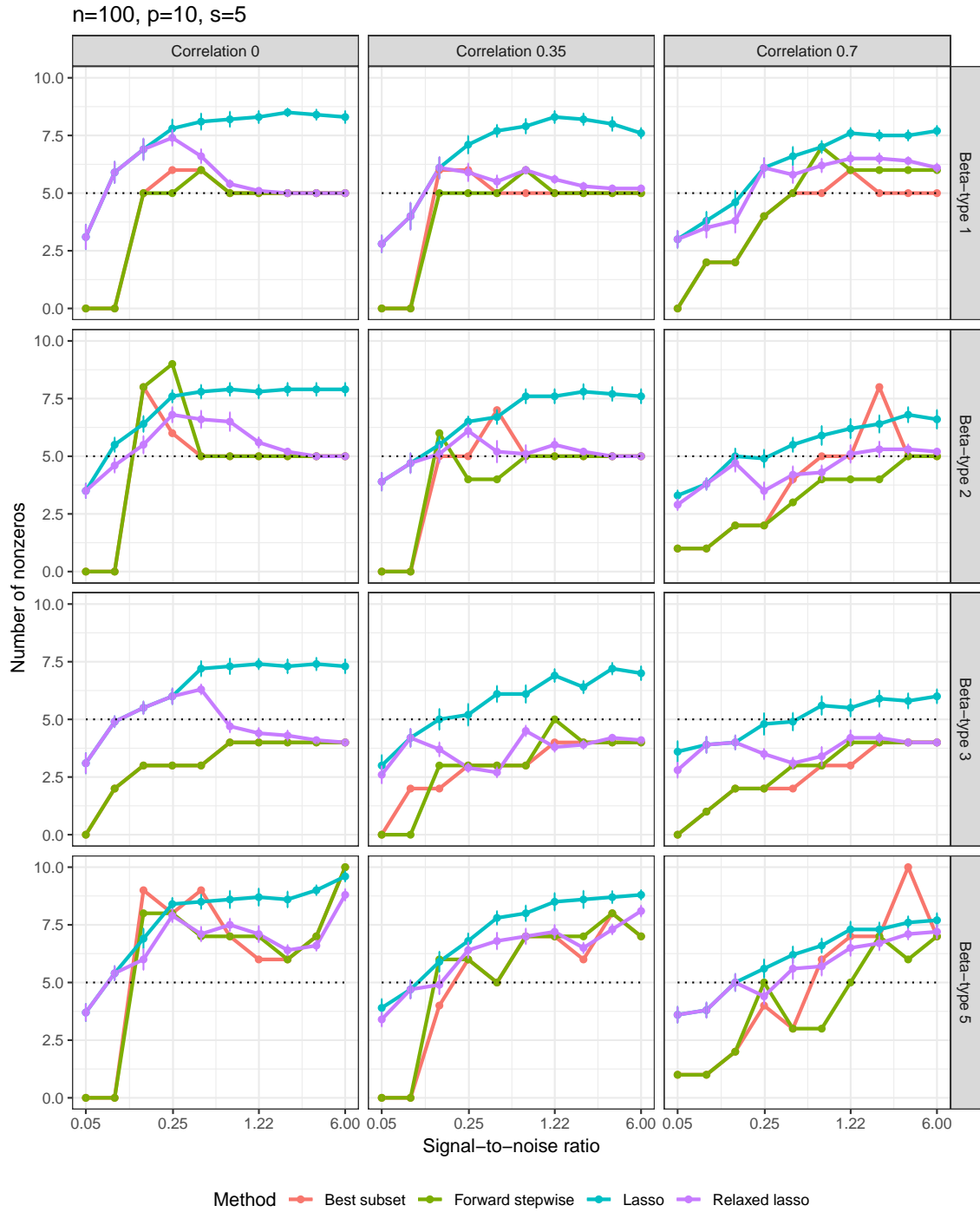
B.1.1 Relative test error (to Bayes)



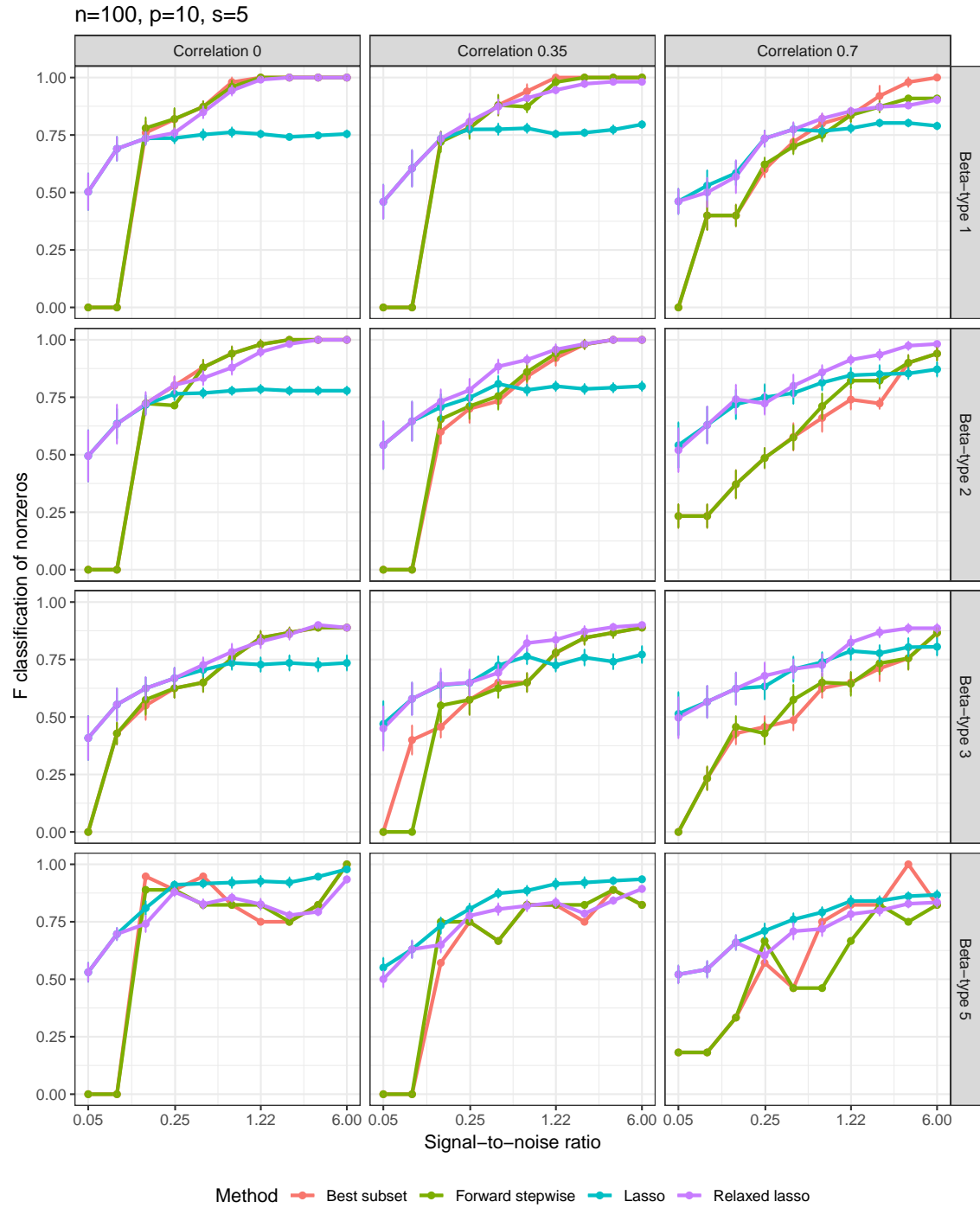
B.1.2 Proportion of variance explained



B.1.3 Number of nonzero coefficients

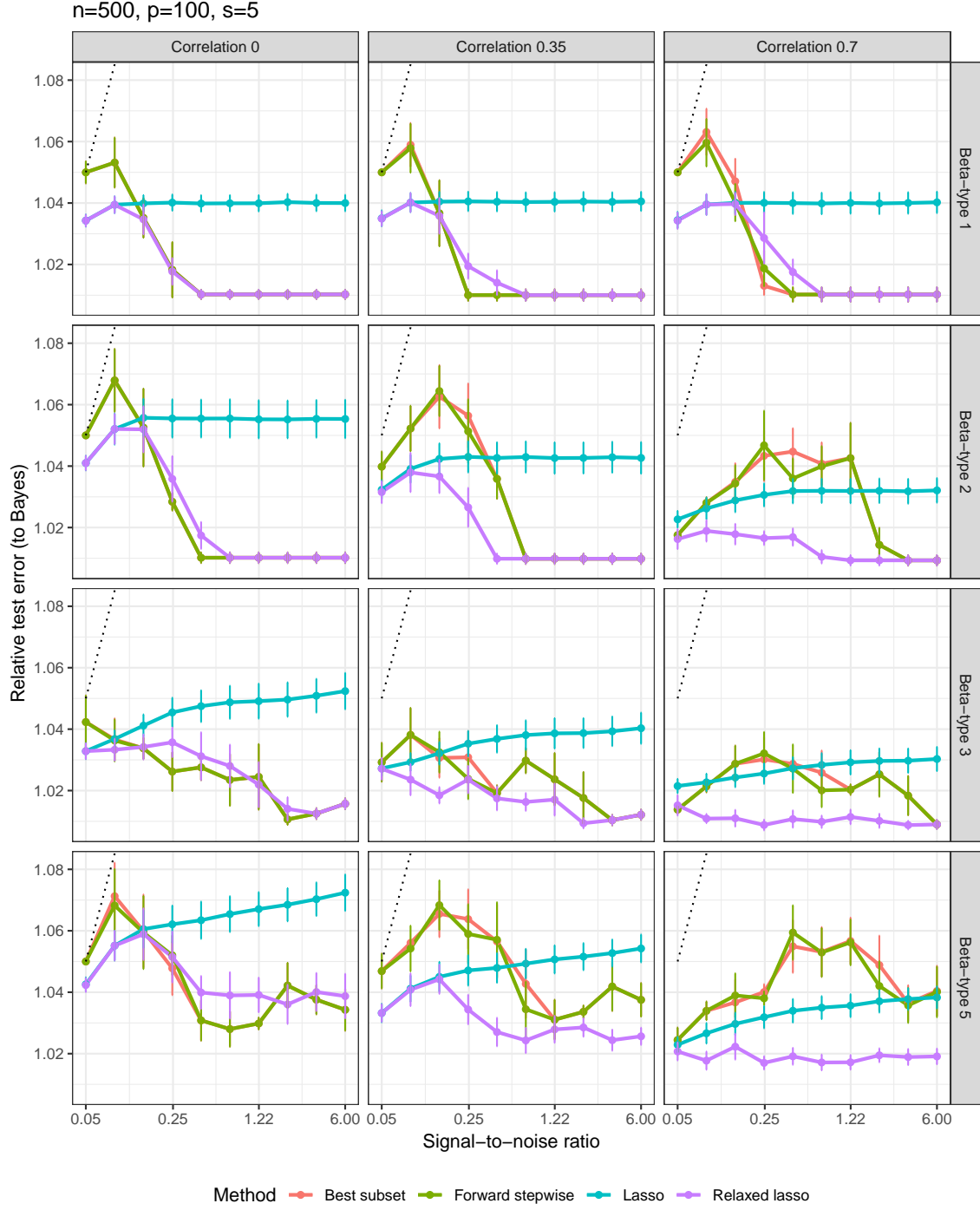


B.1.4 F-score

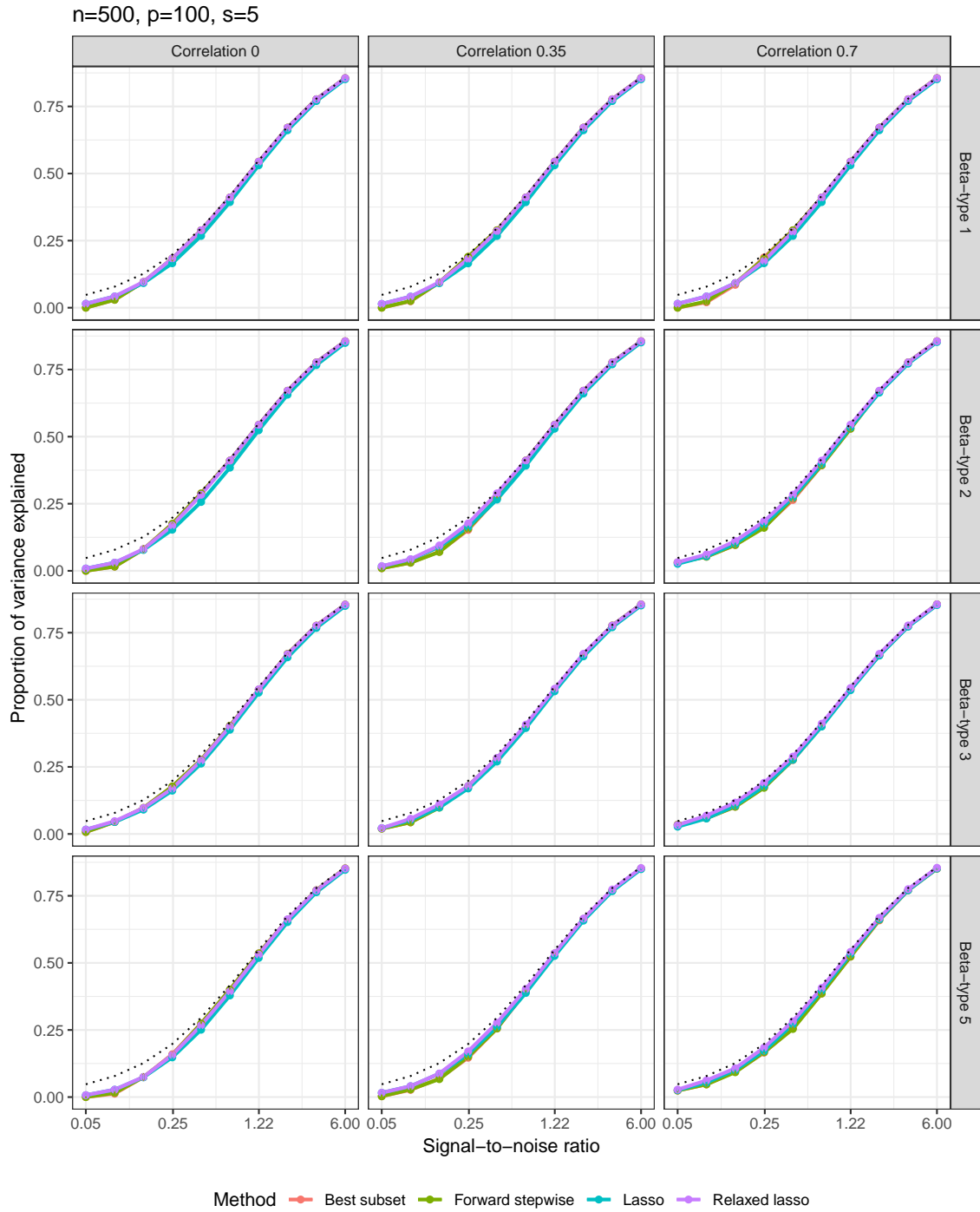


B.2 Medium setting: $n = 500, p = 100, s = 5$

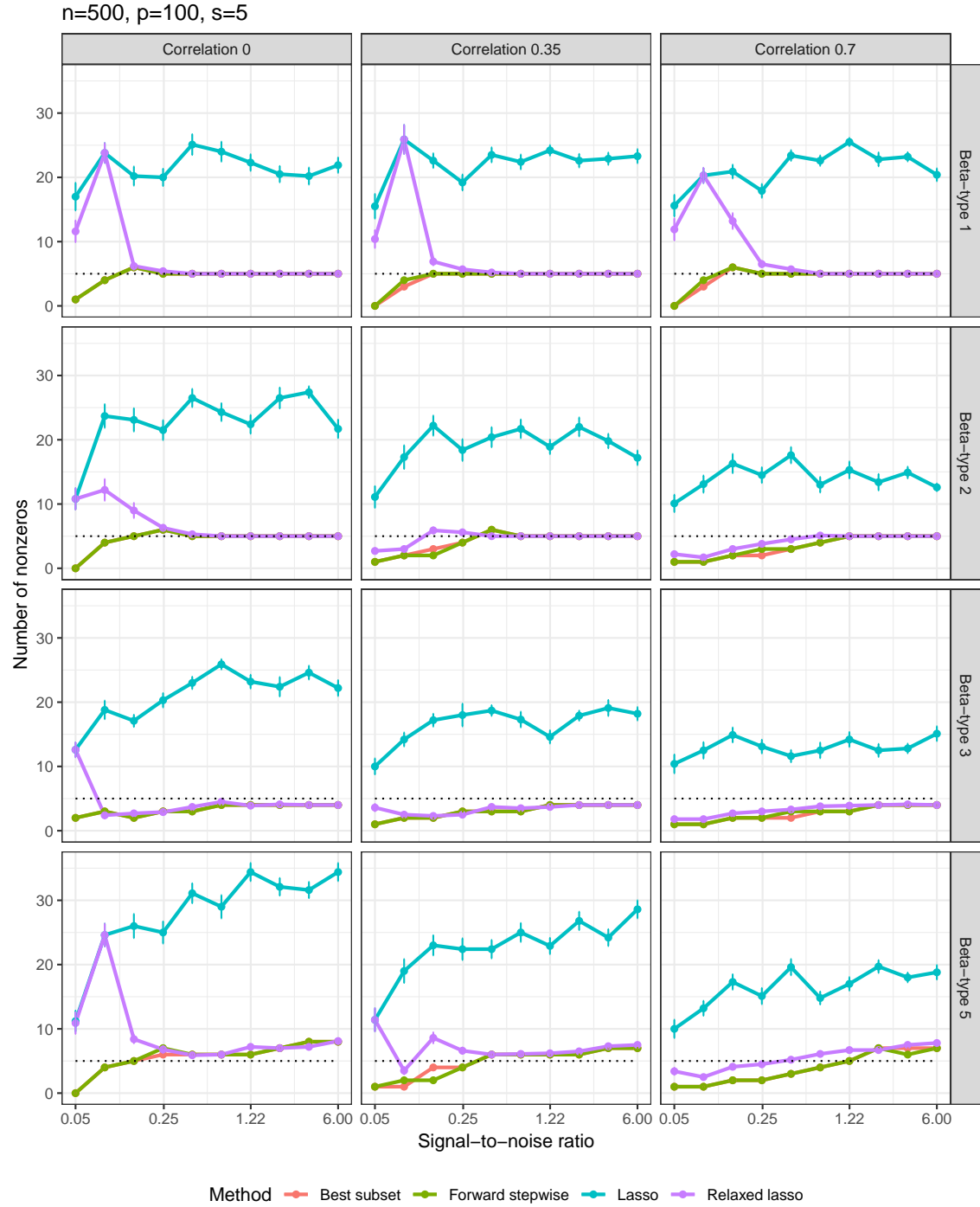
B.2.1 Relative test error (to Bayes)



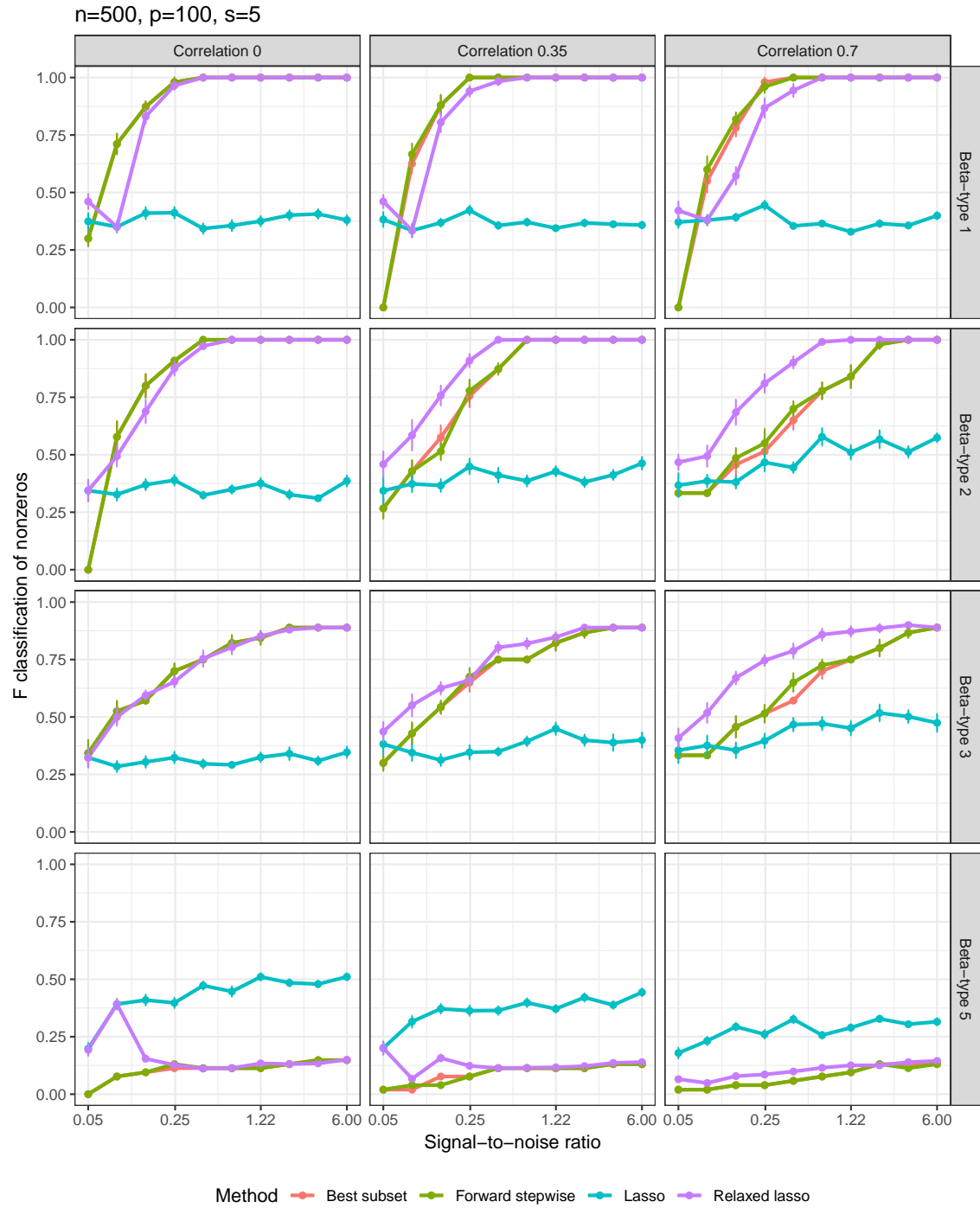
B.2.2 Proportion of variance explained



B.2.3 Number of nonzero coefficients

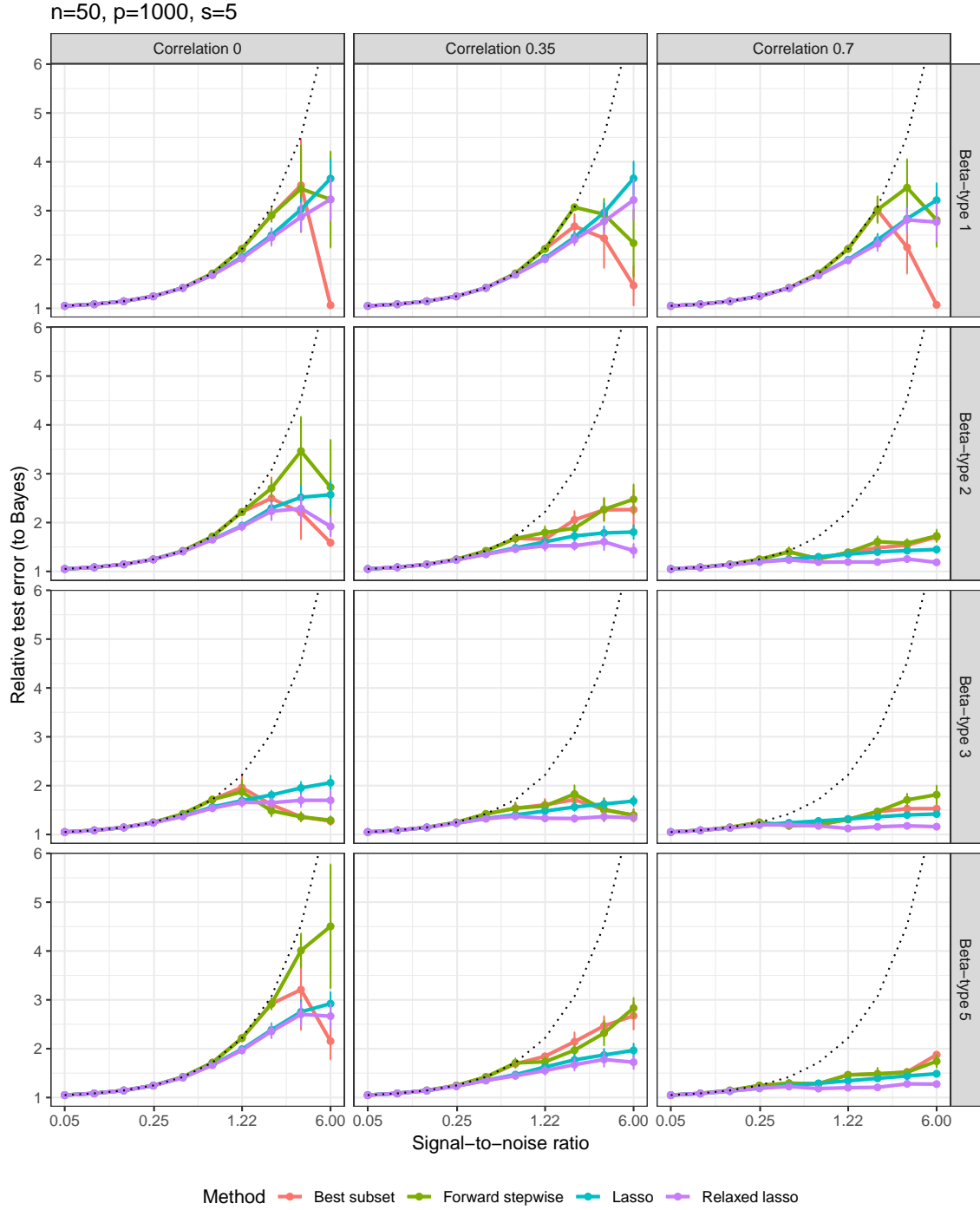


B.2.4 F-score

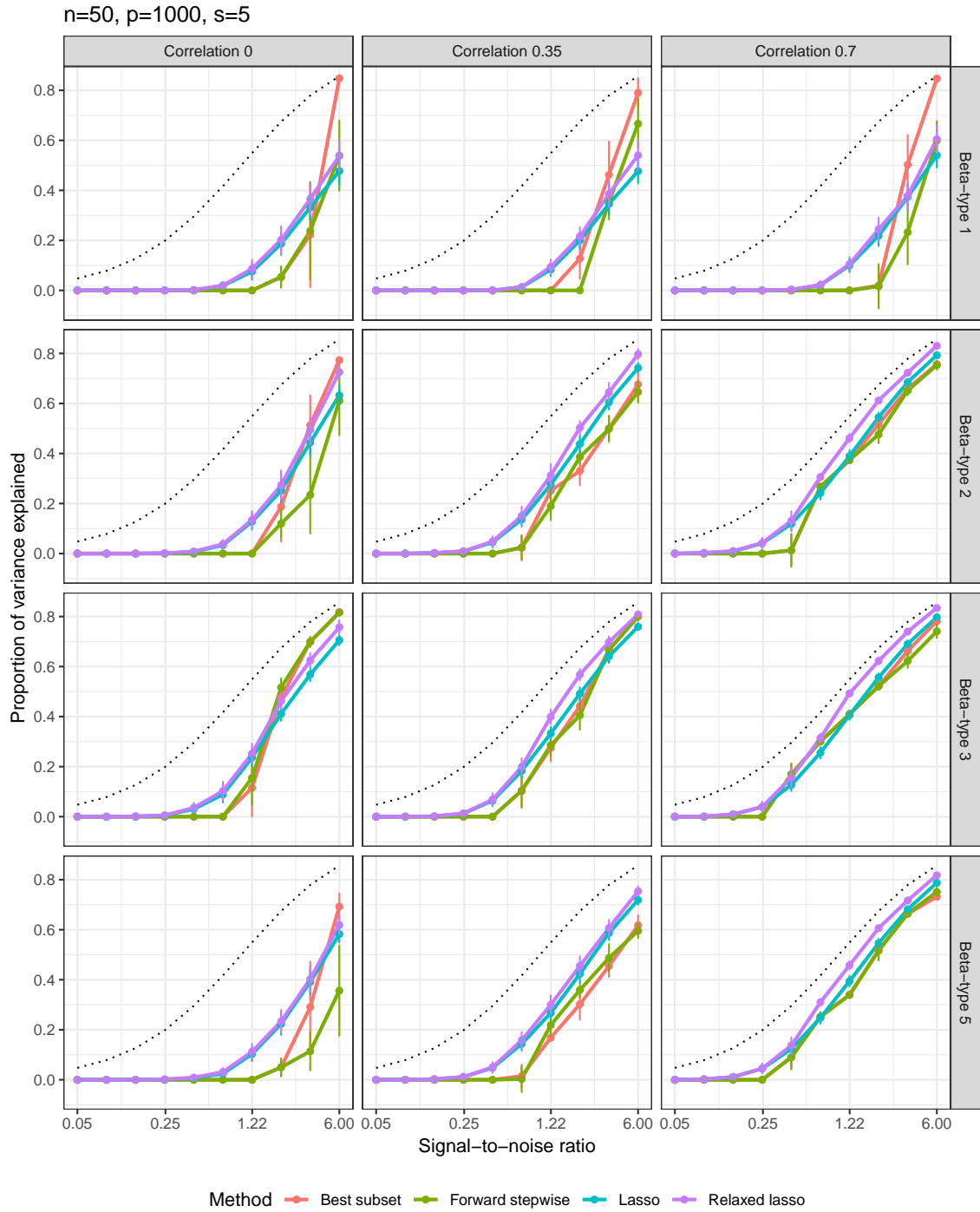


B.3 High-5 setting: $n = 50$, $p = 1000$, $s = 5$

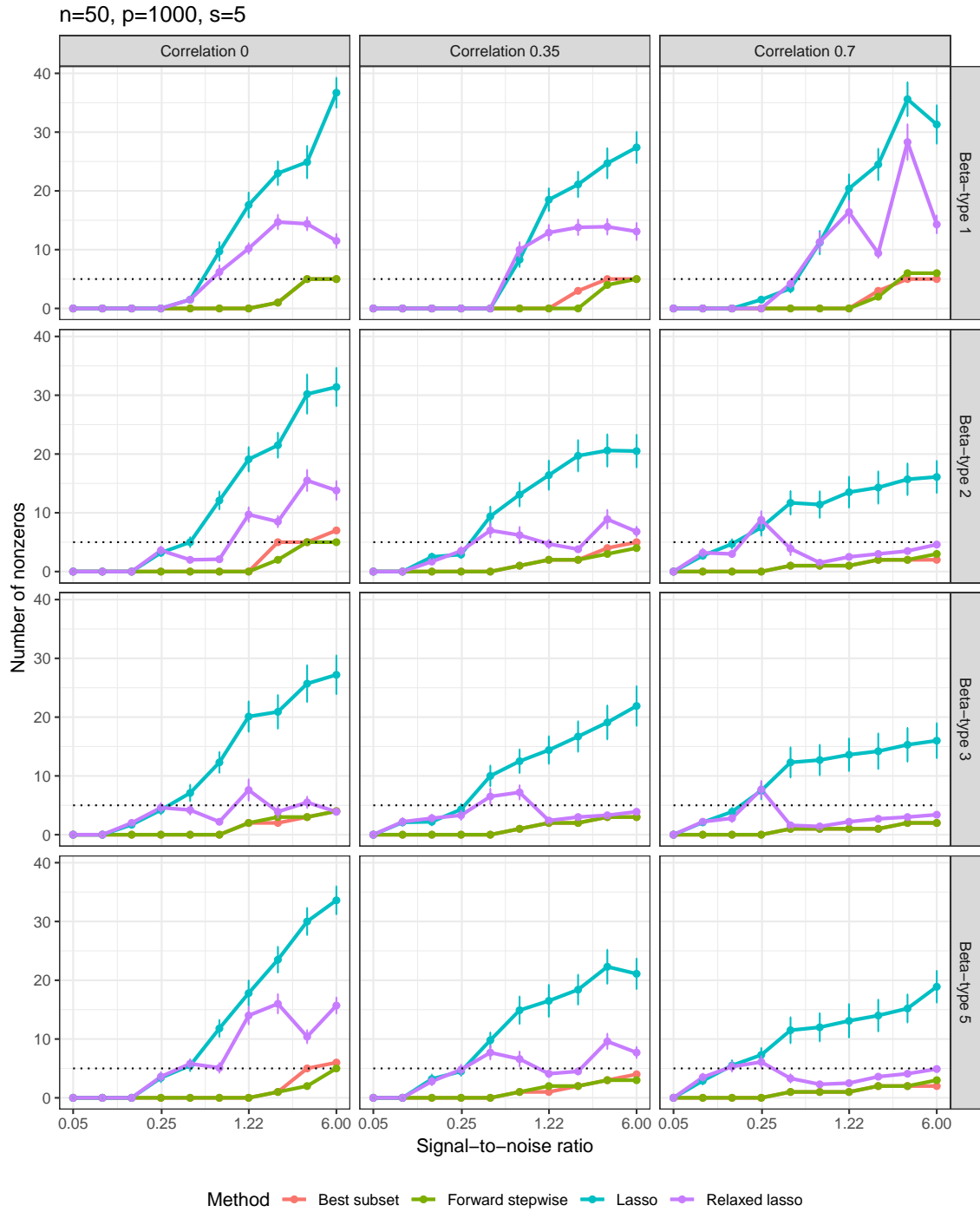
B.3.1 Relative test error (to Bayes)



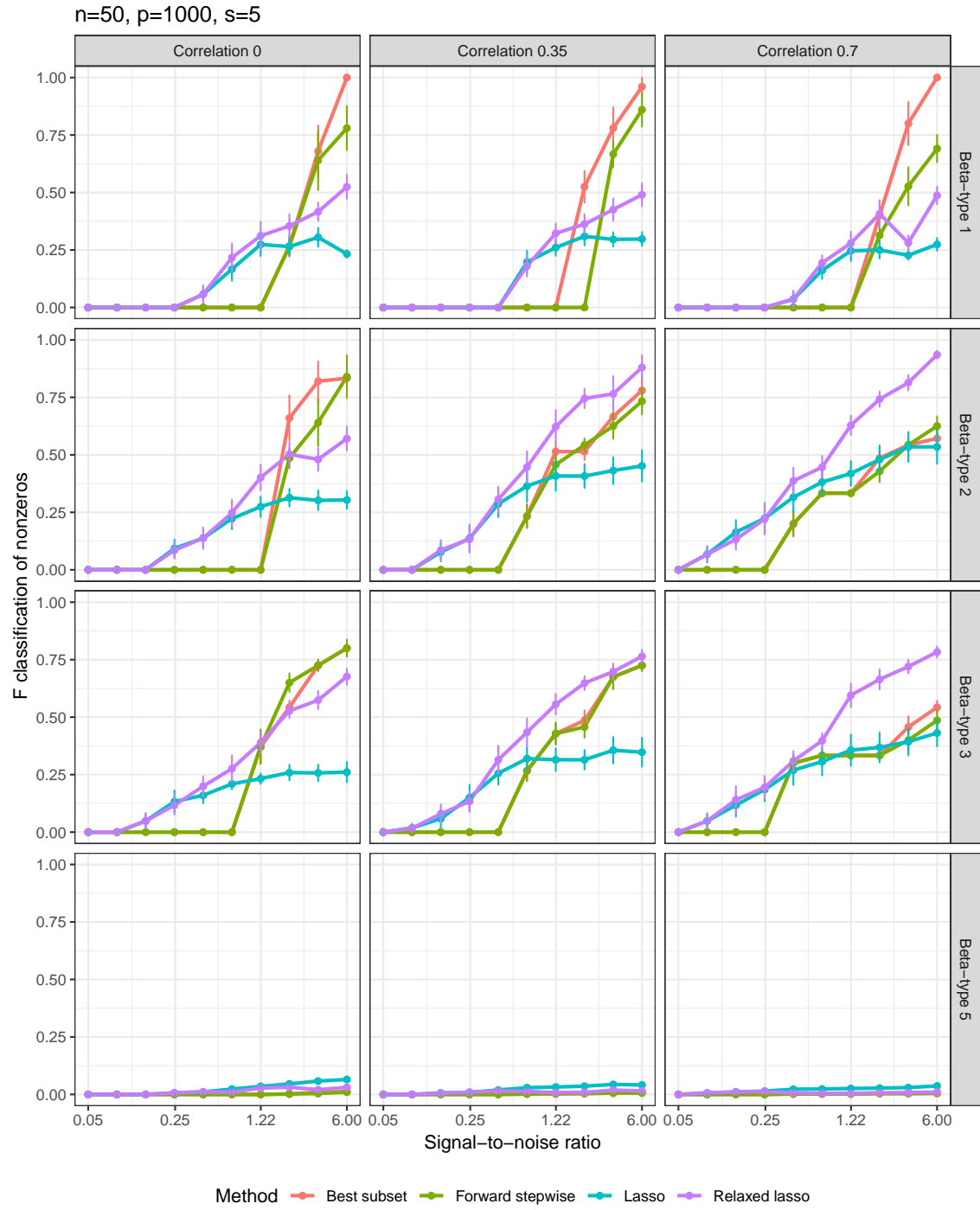
B.3.2 Proportion of variance explained



B.3.3 Number of nonzero coefficients

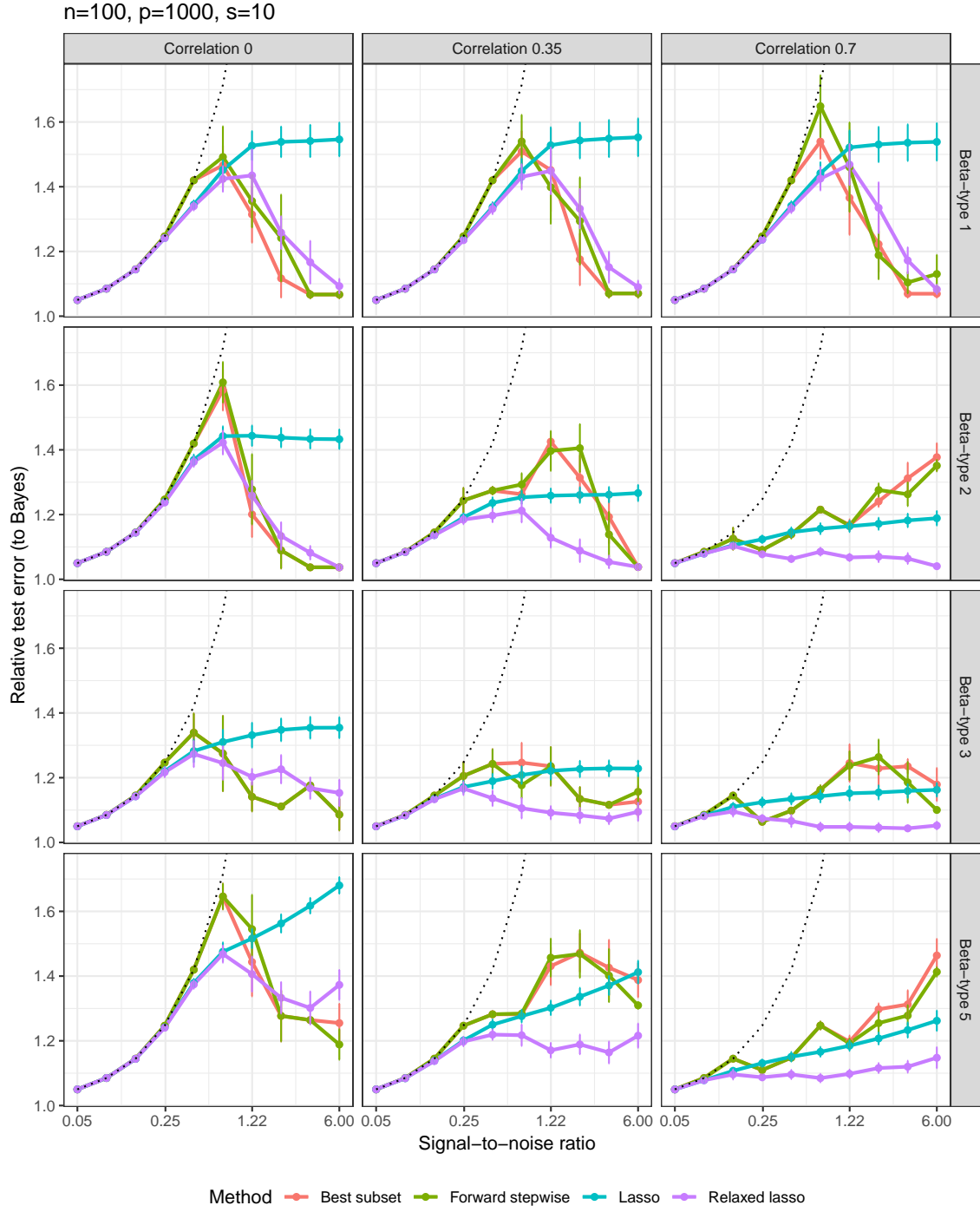


B.3.4 F-score

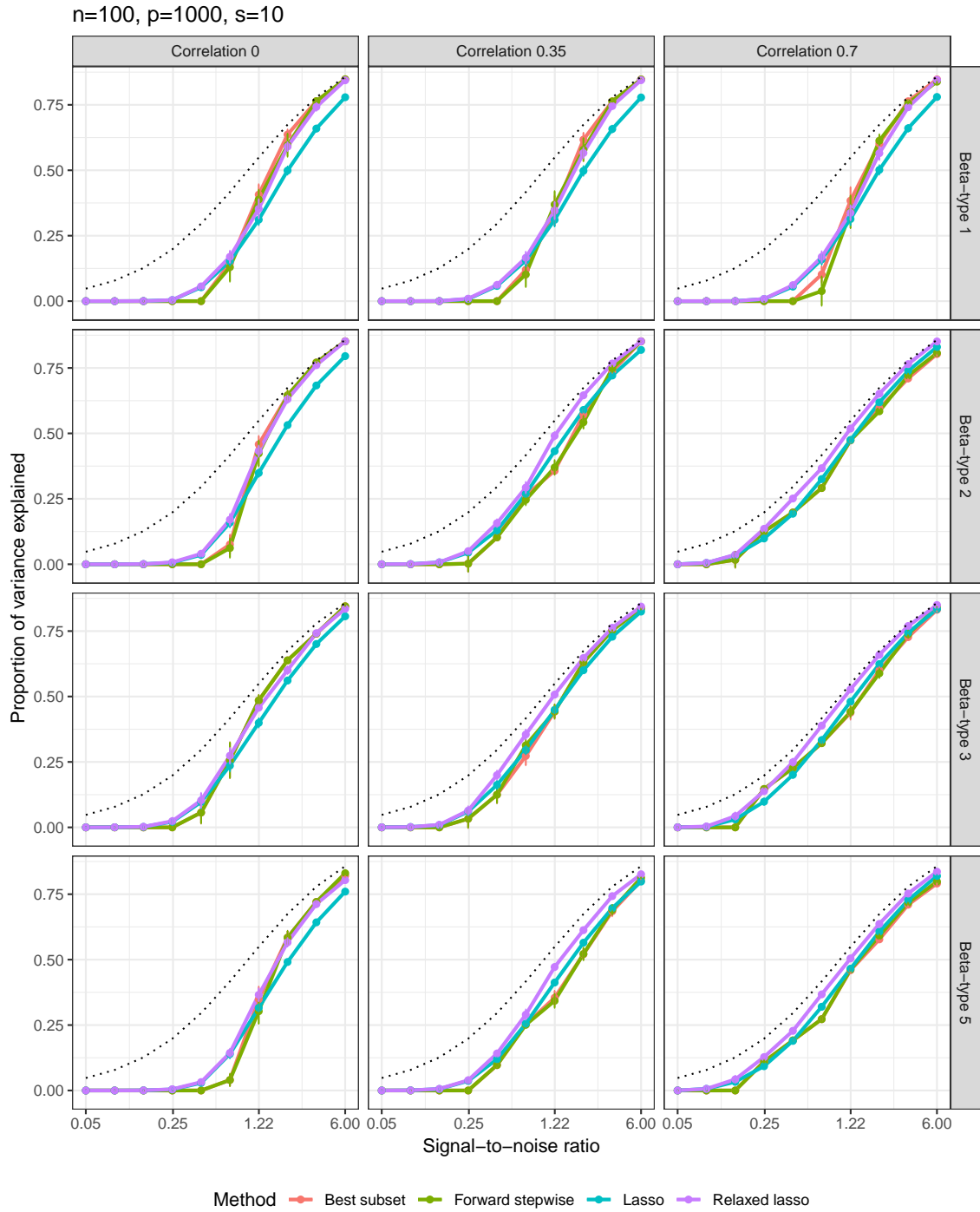


B.4 High-10 setting: $n = 100$, $p = 1000$, $s = 10$

B.4.1 Relative test error (to Bayes)

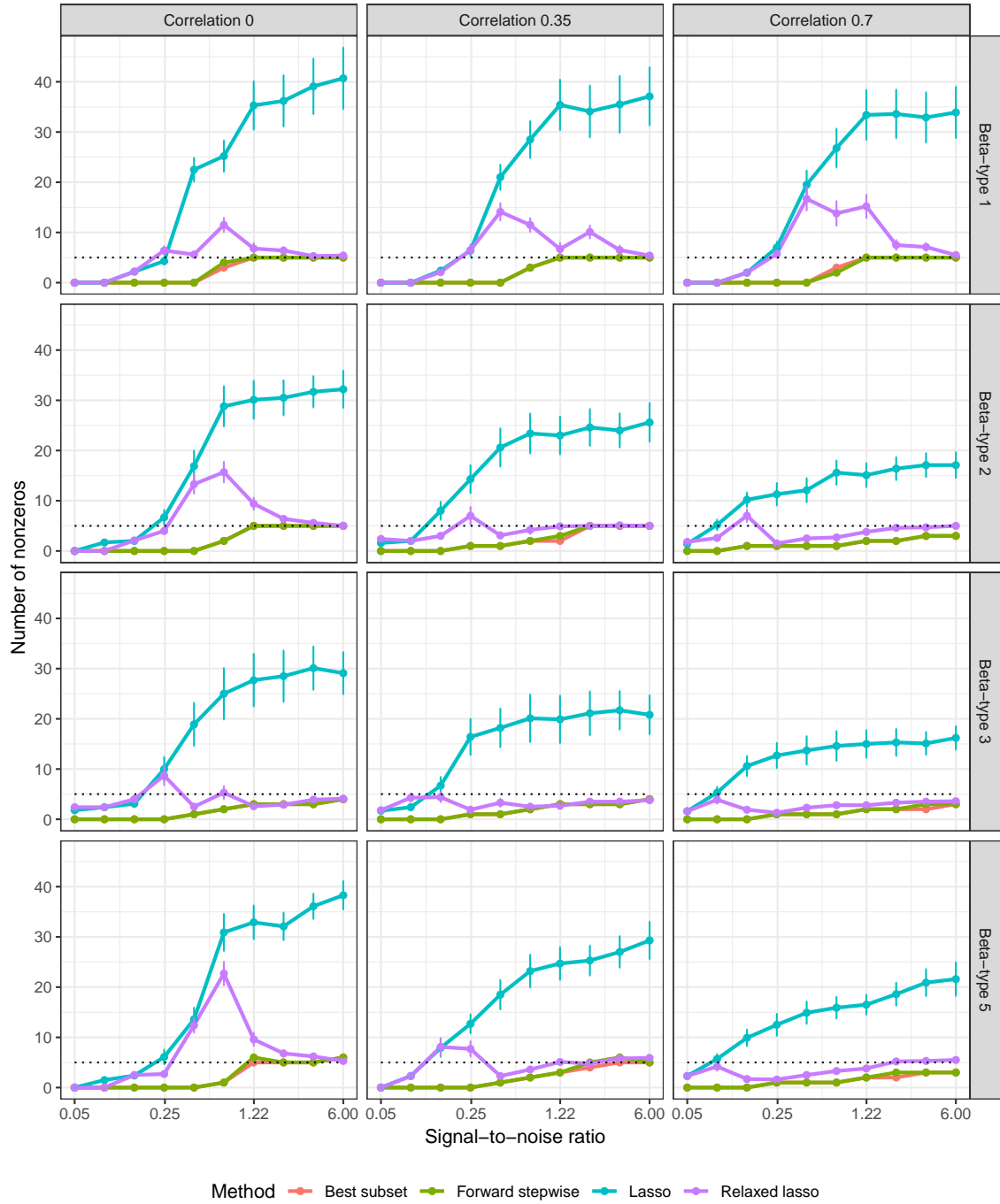


B.4.2 Proportion of variance explained

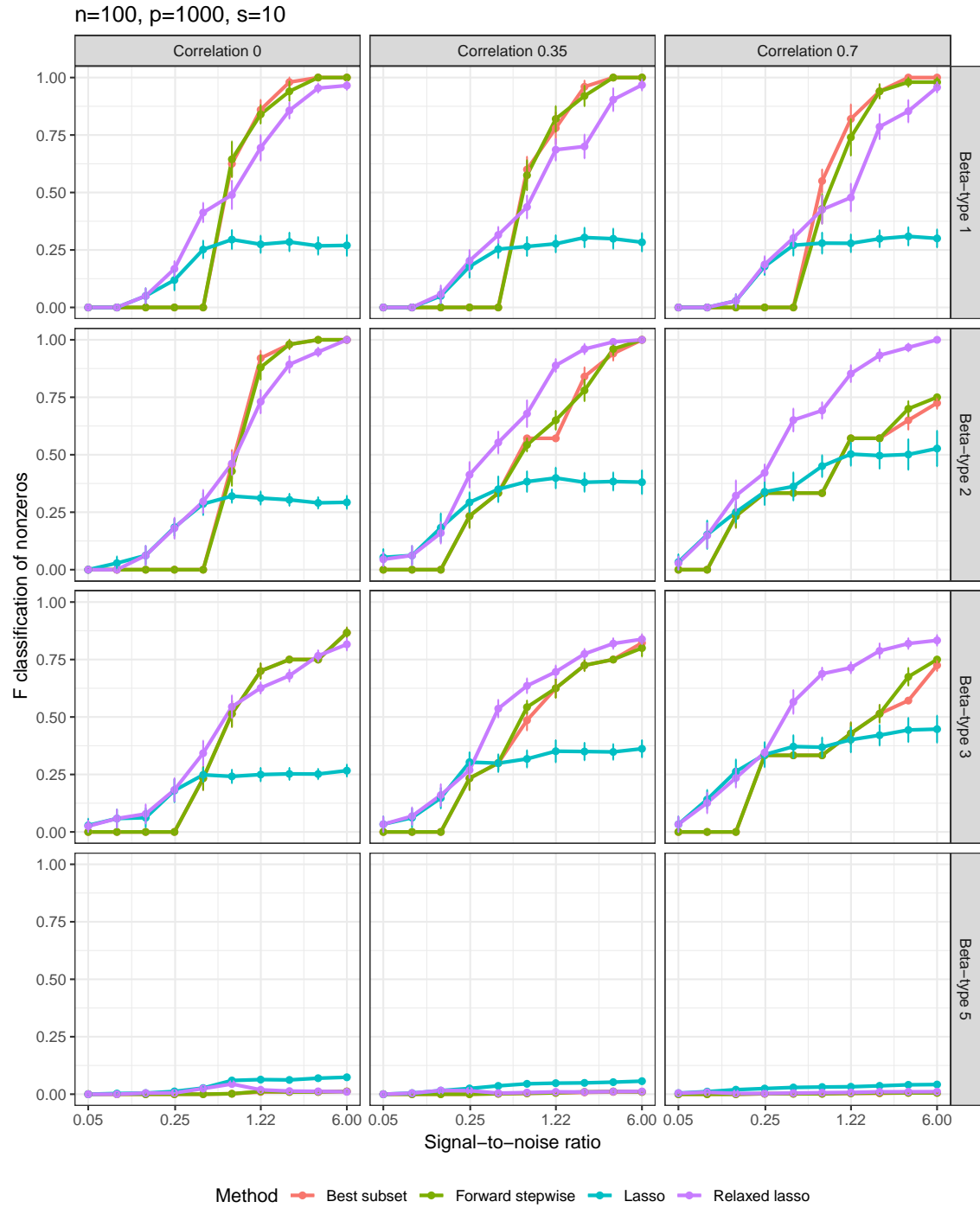


B.4.3 Number of nonzero coefficients

$n=100, p=1000, s=10$



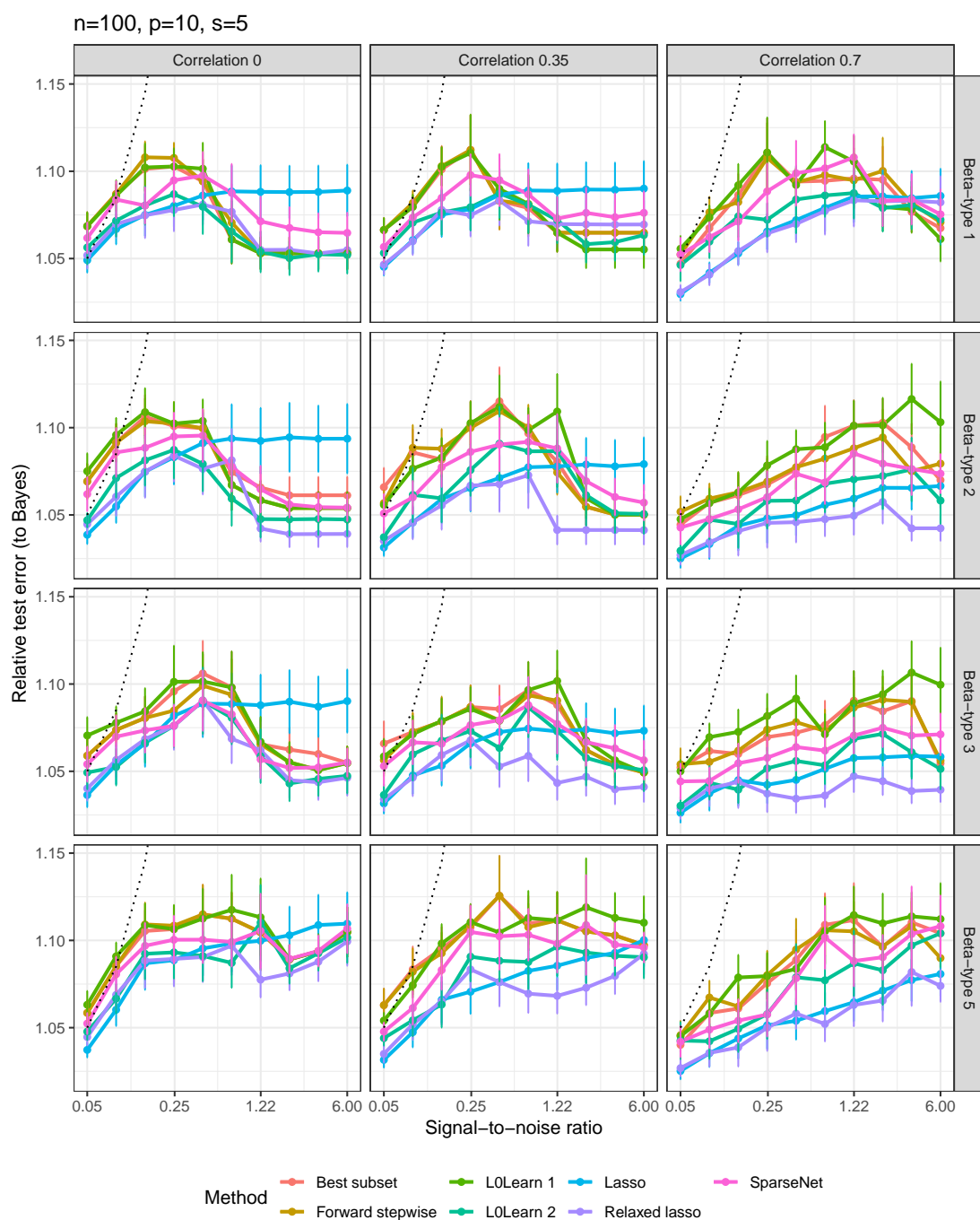
B.4.4 F-score



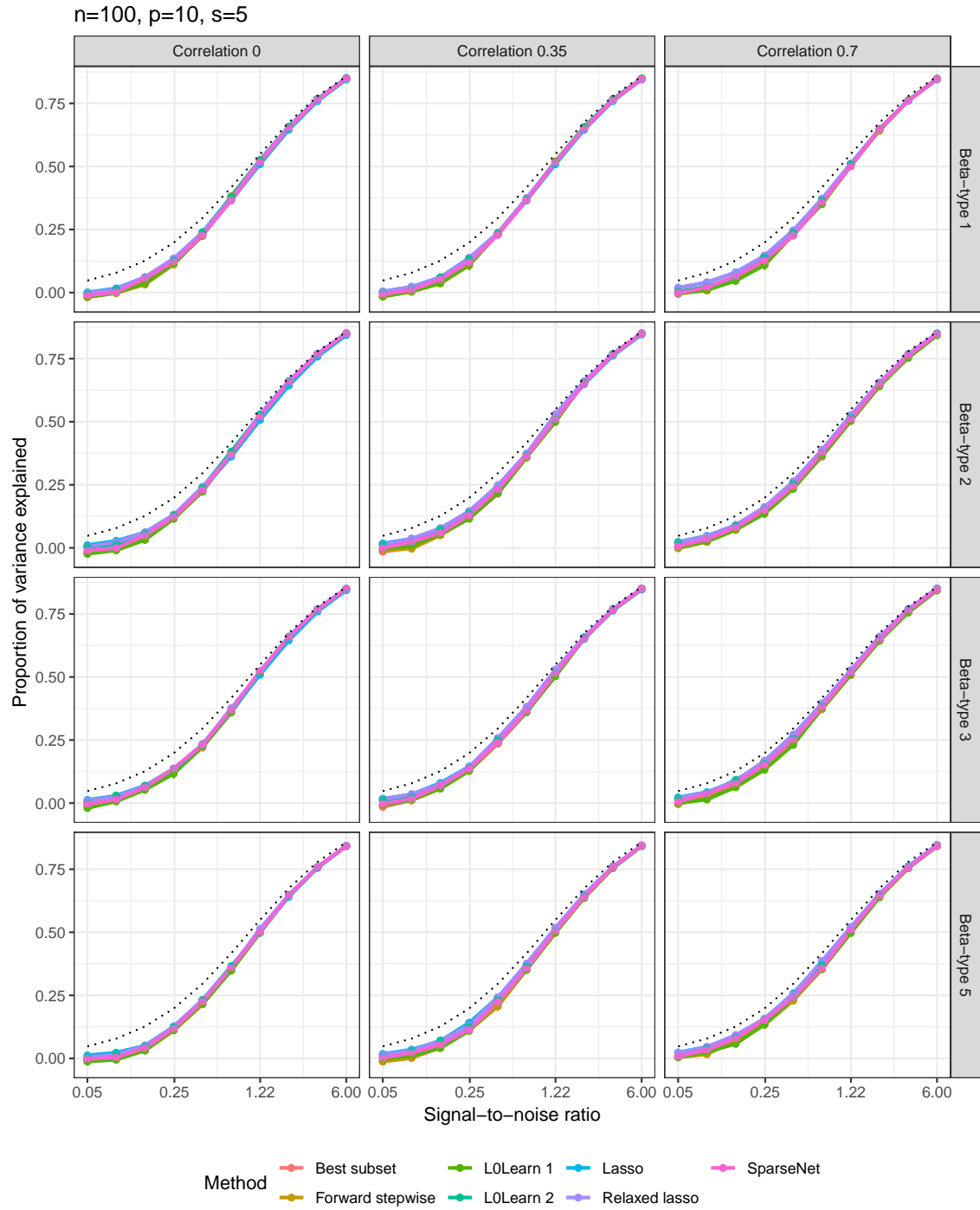
C Validation tuning: now including L0Learn-1, L0Learn-2 and SparseNet

C.1 Low setting: $n = 100$, $p = 10$, $s = 5$

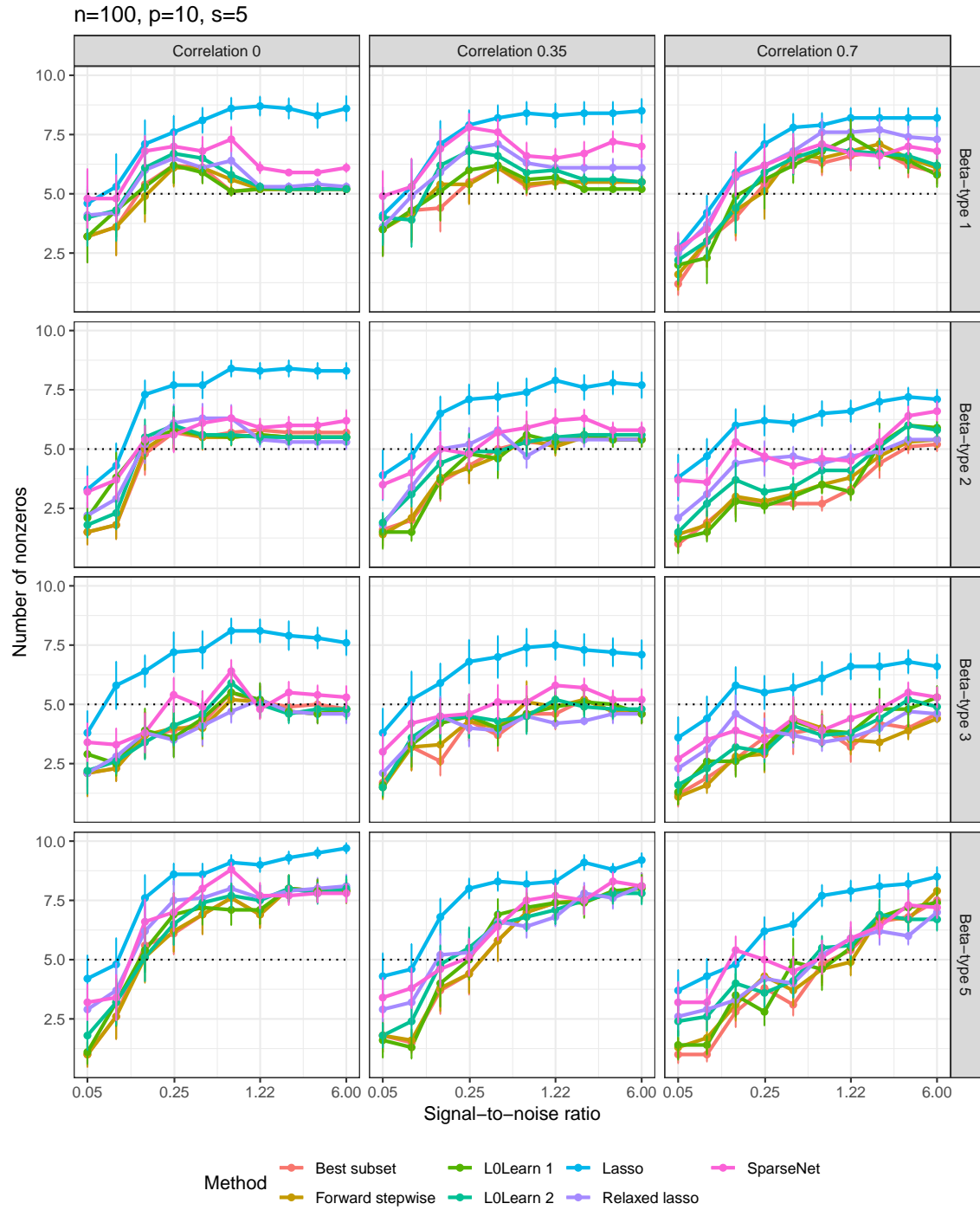
C.1.1 Relative test error (to Bayes)



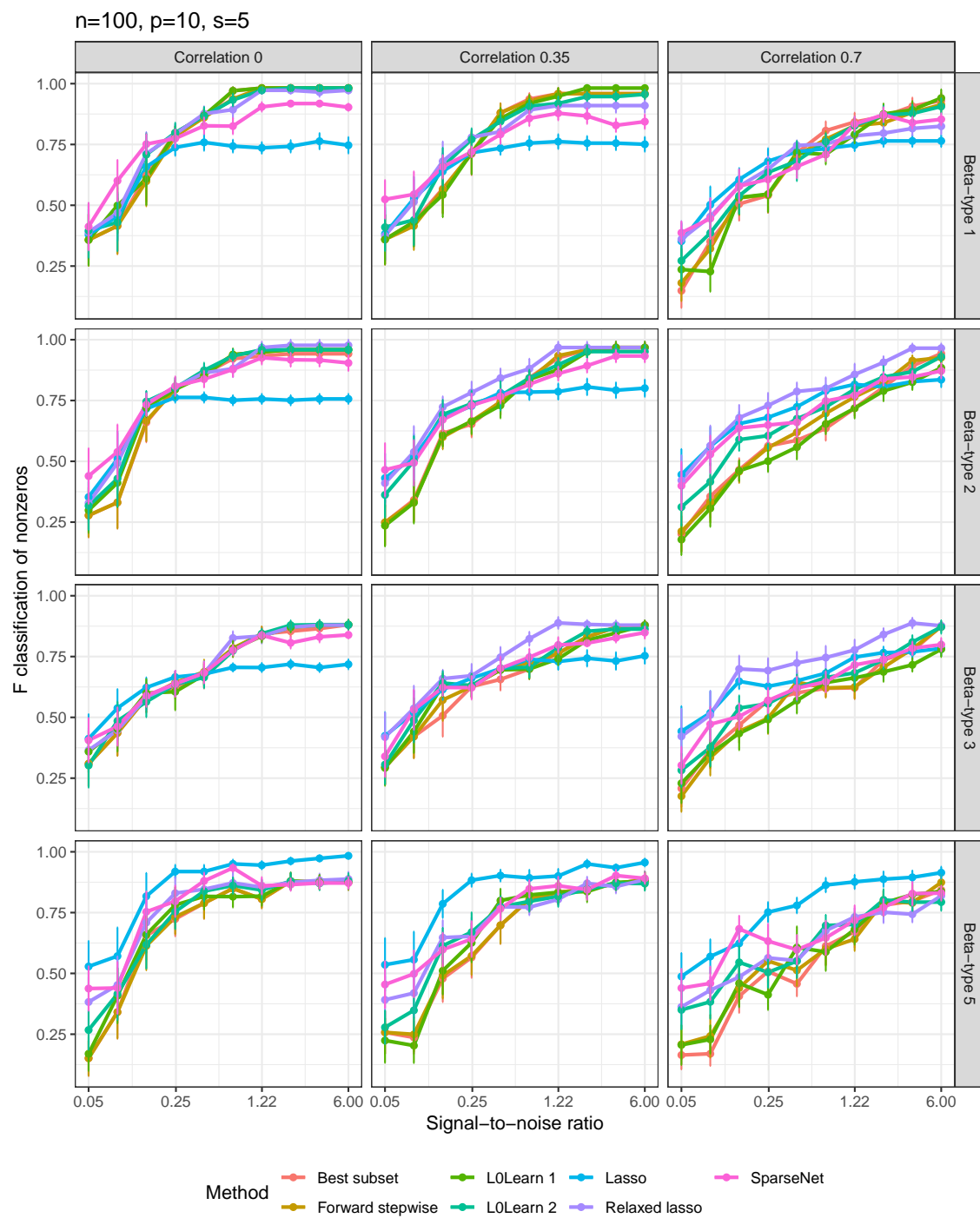
C.1.2 Proportion of variance explained



C.1.3 Number of nonzero coefficients

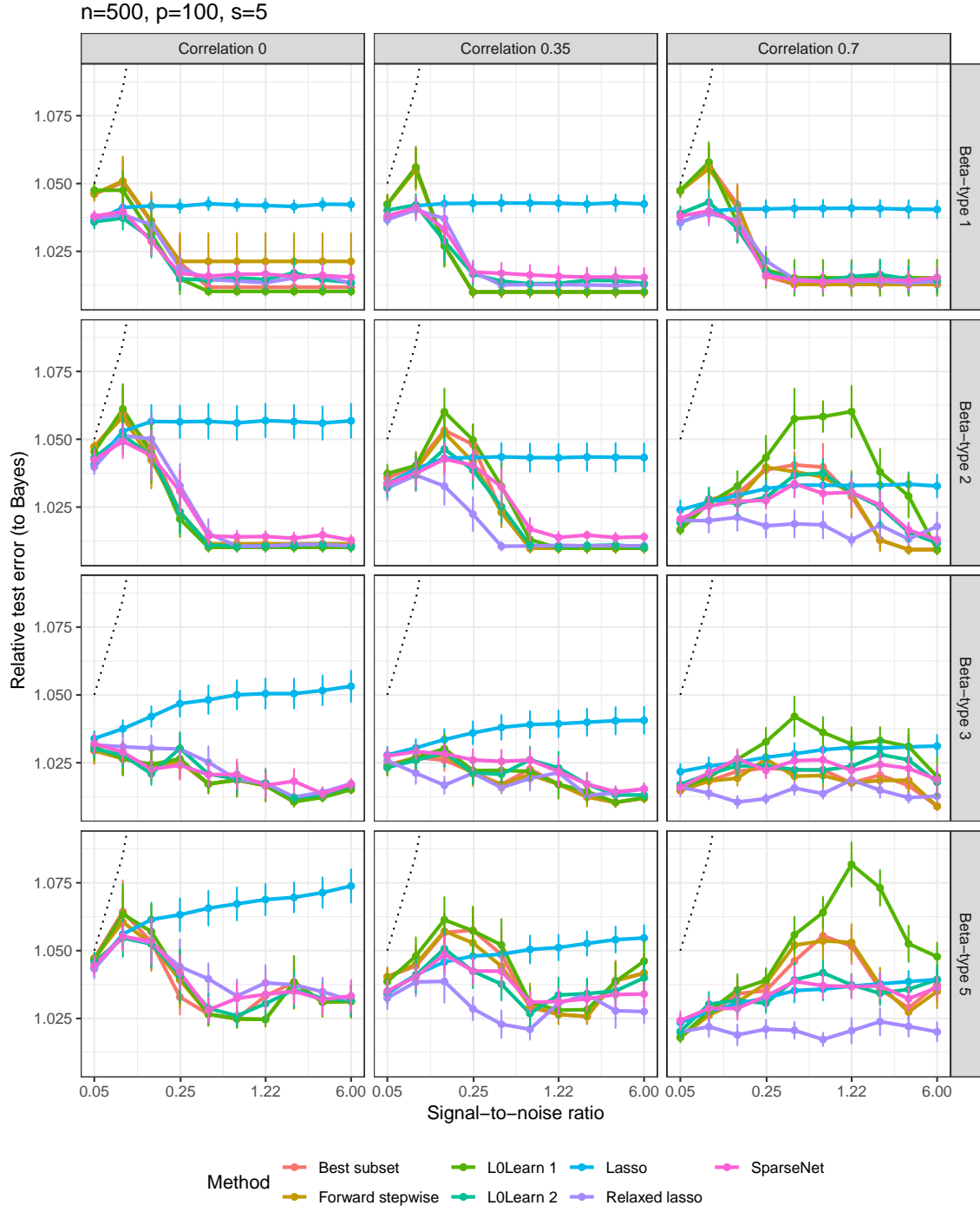


C.1.4 F-score

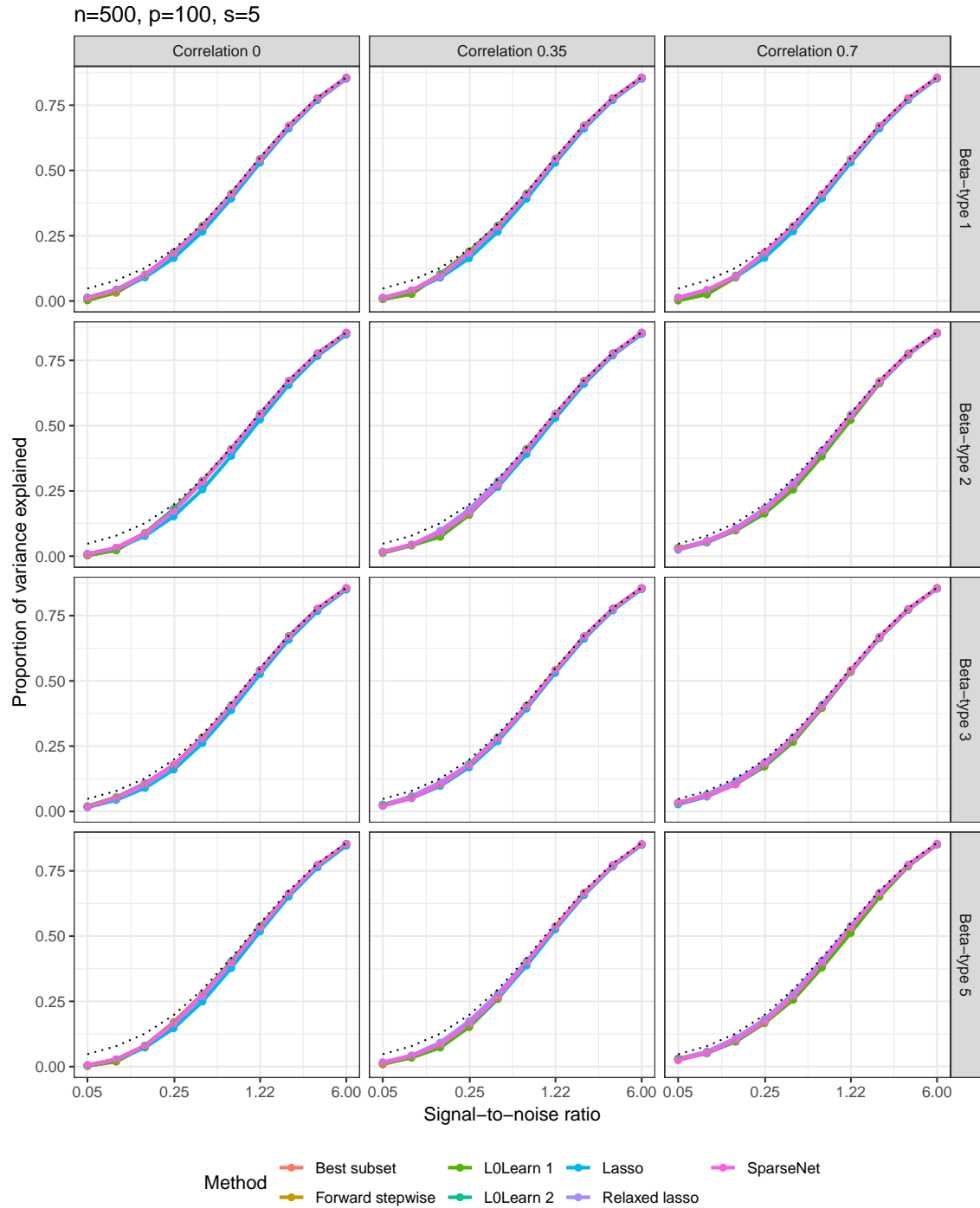


C.2 Medium setting: $n = 500, p = 100, s = 5$

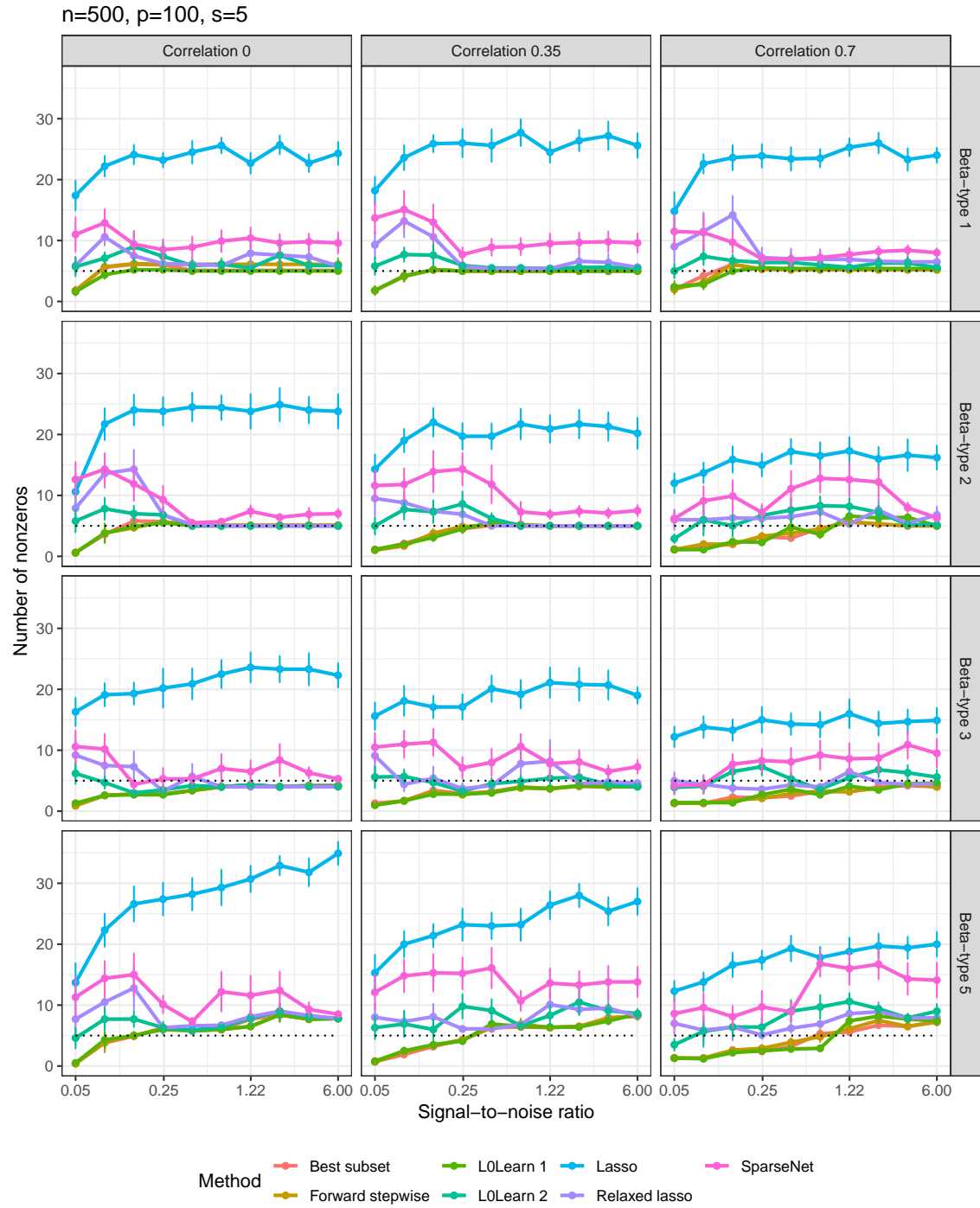
C.2.1 Relative test error (to Bayes)



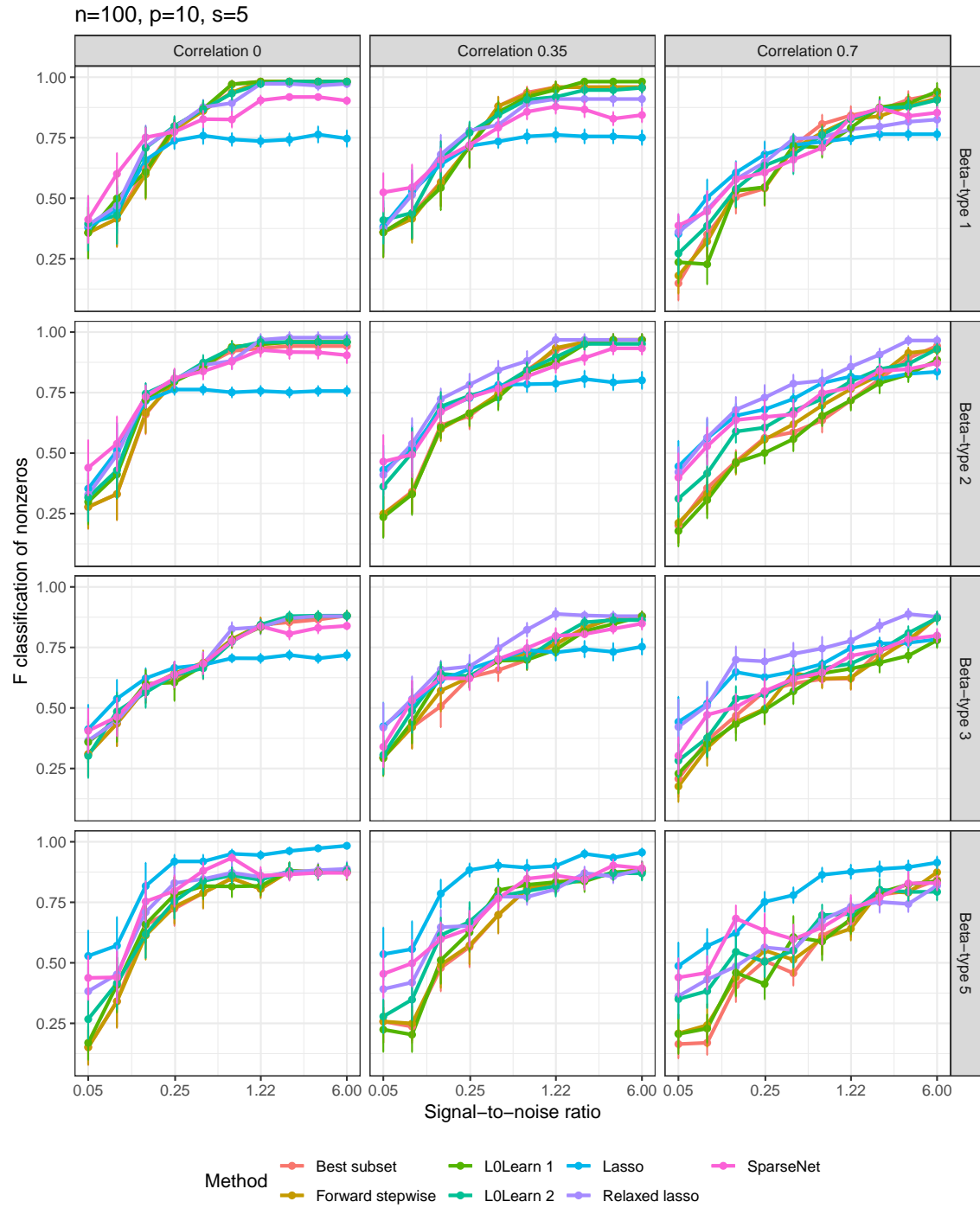
C.2.2 Proportion of variance explained



C.2.3 Number of nonzero coefficients

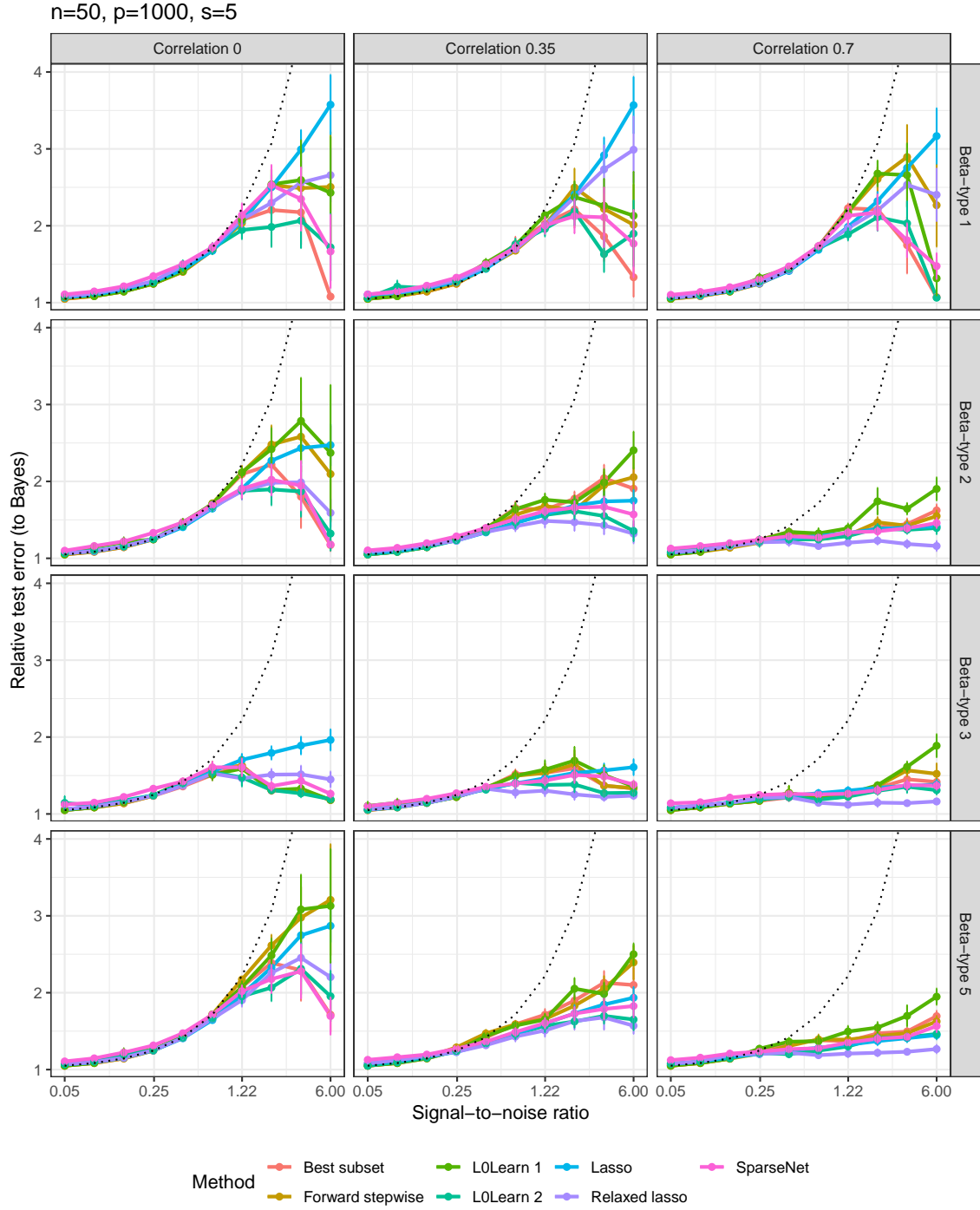


C.2.4 F-score

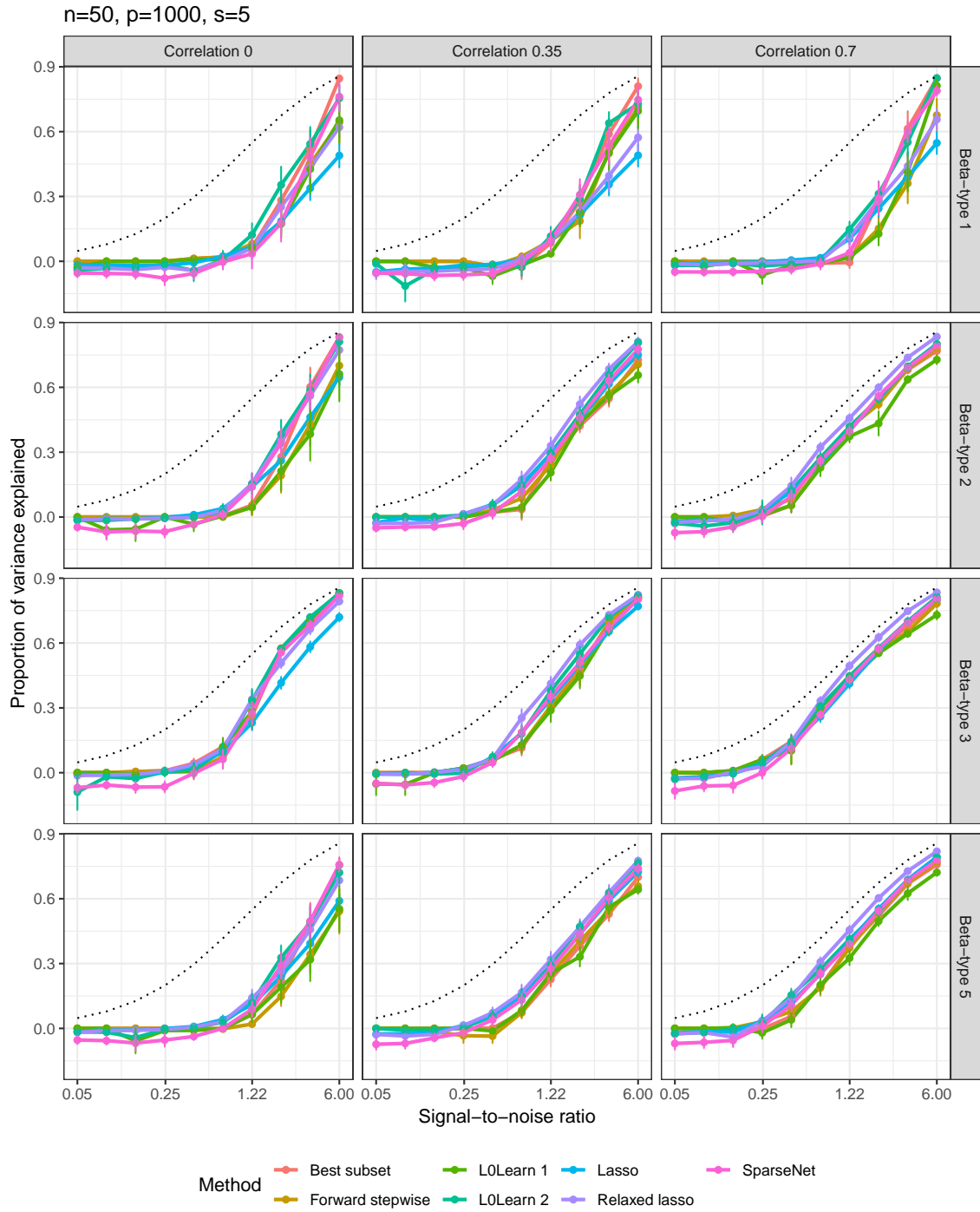


C.3 High-5 setting: $n = 50$, $p = 1000$, $s = 5$

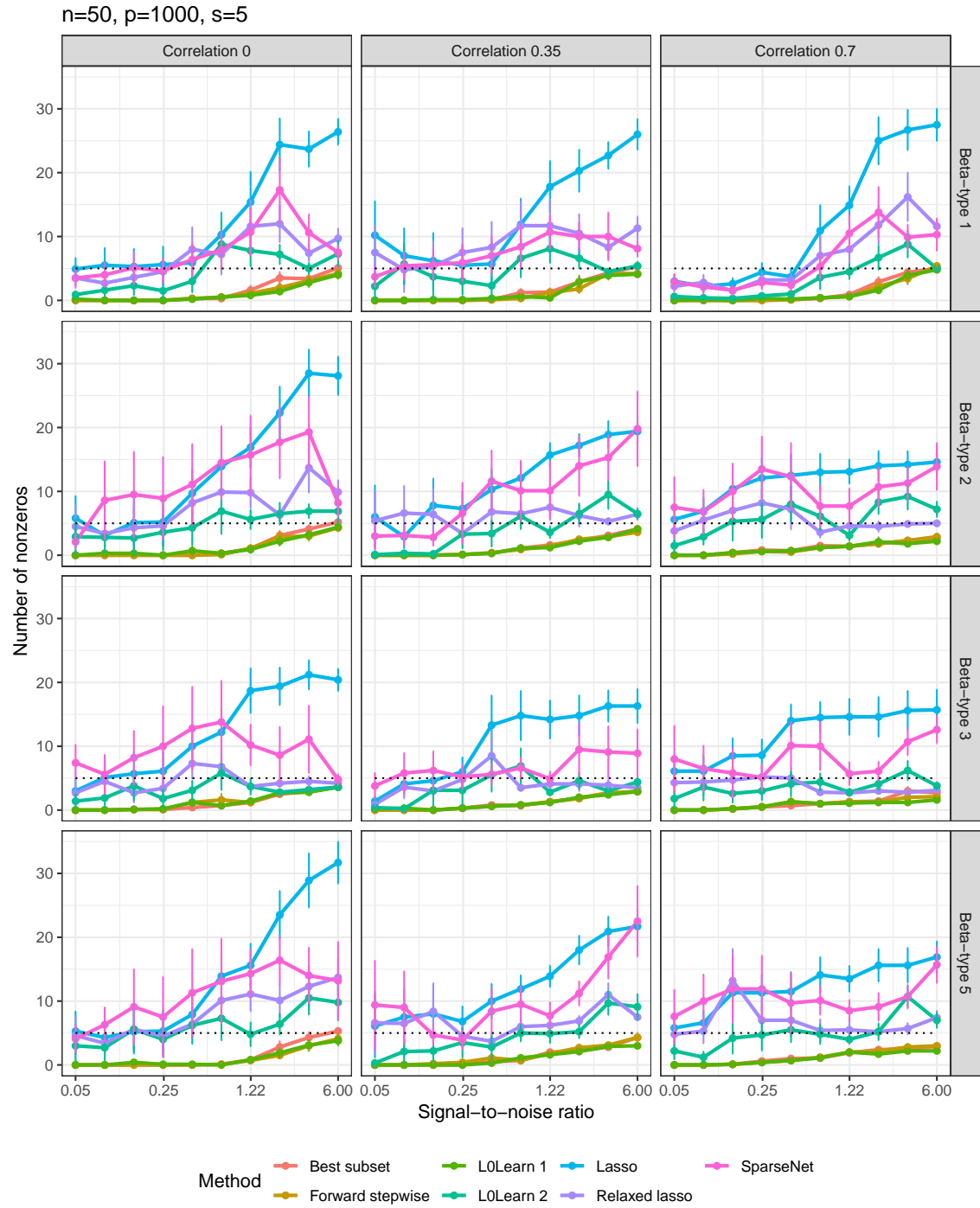
C.3.1 Relative test error (to Bayes)



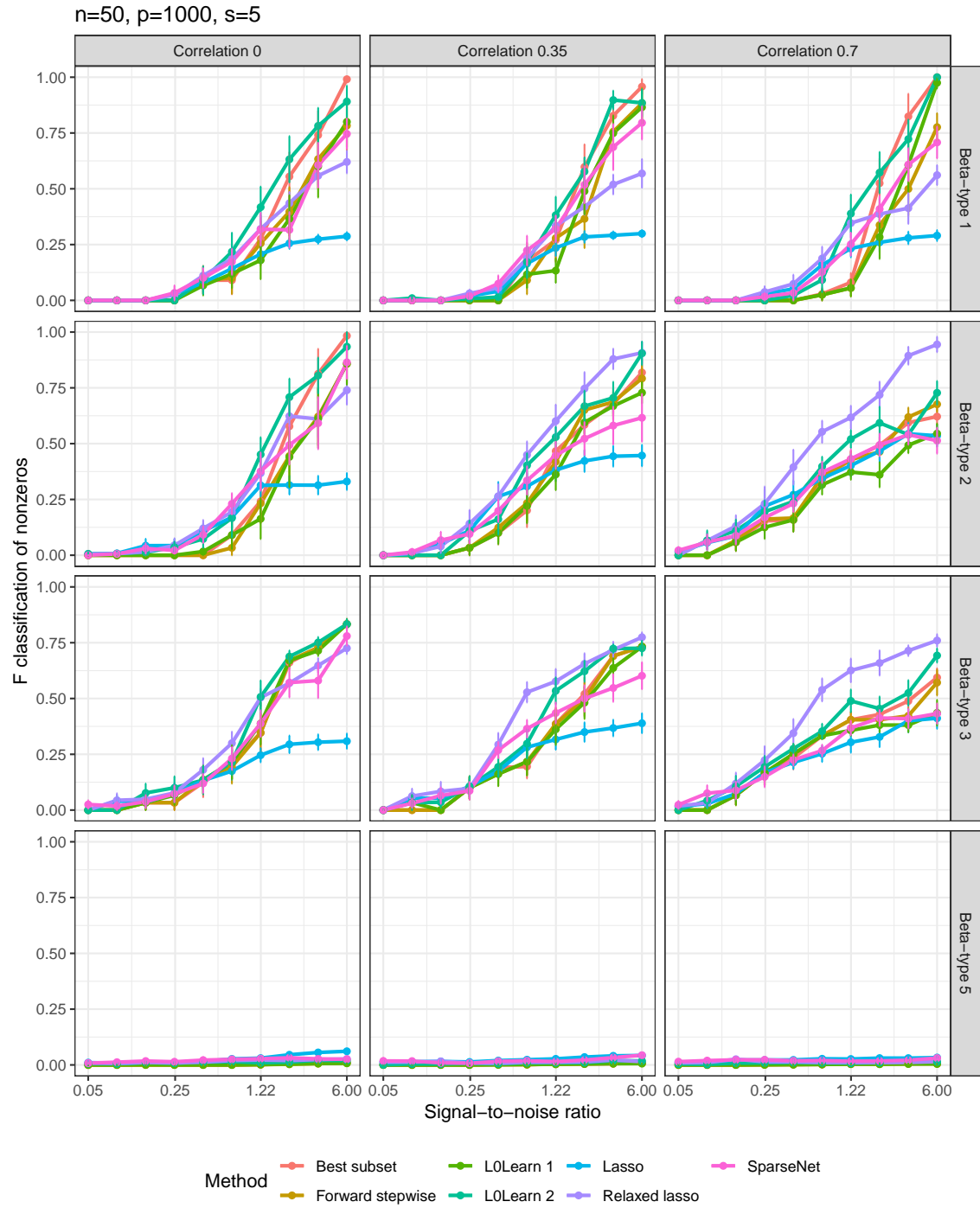
C.3.2 Proportion of variance explained



C.3.3 Number of nonzero coefficients

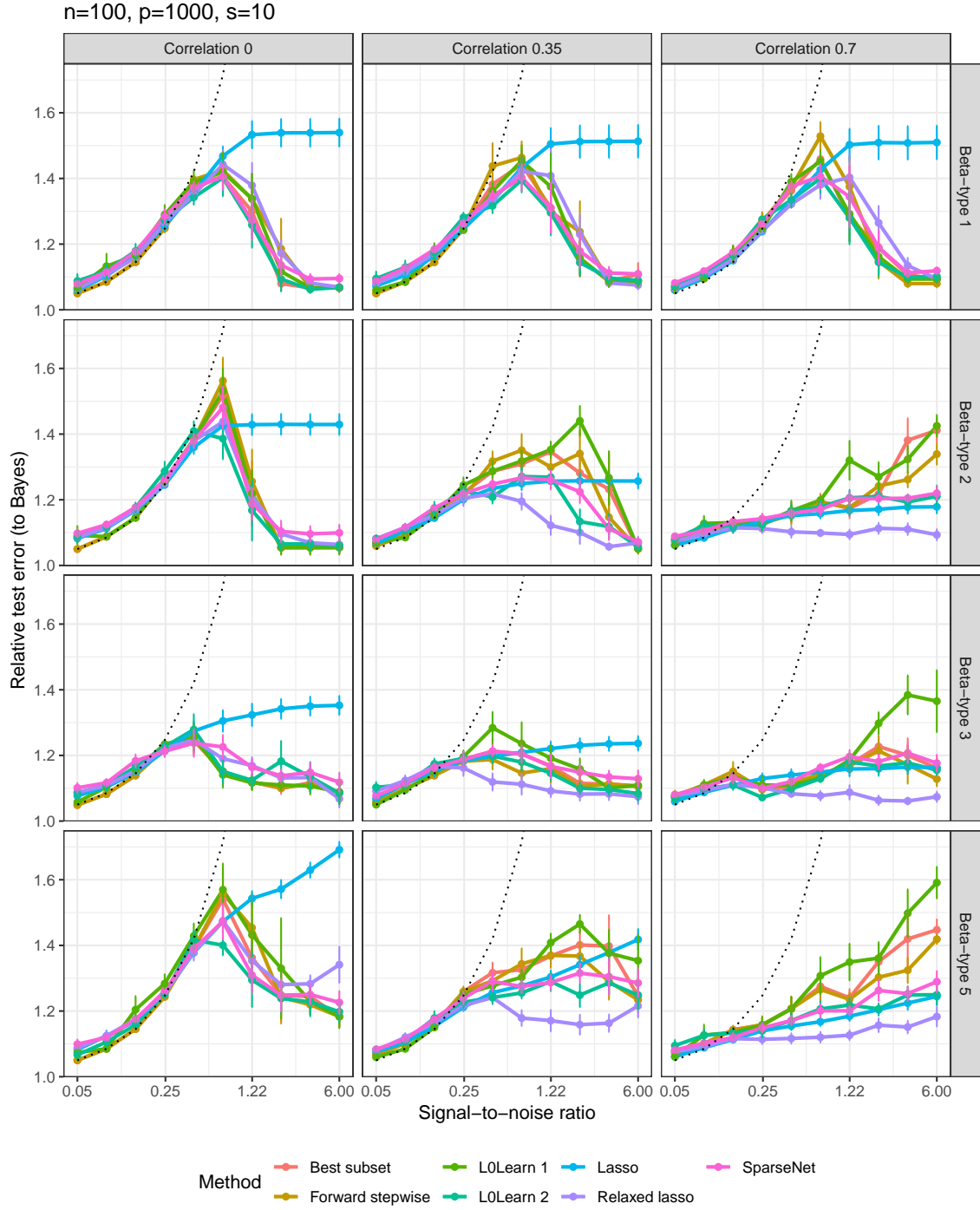


C.3.4 F-score

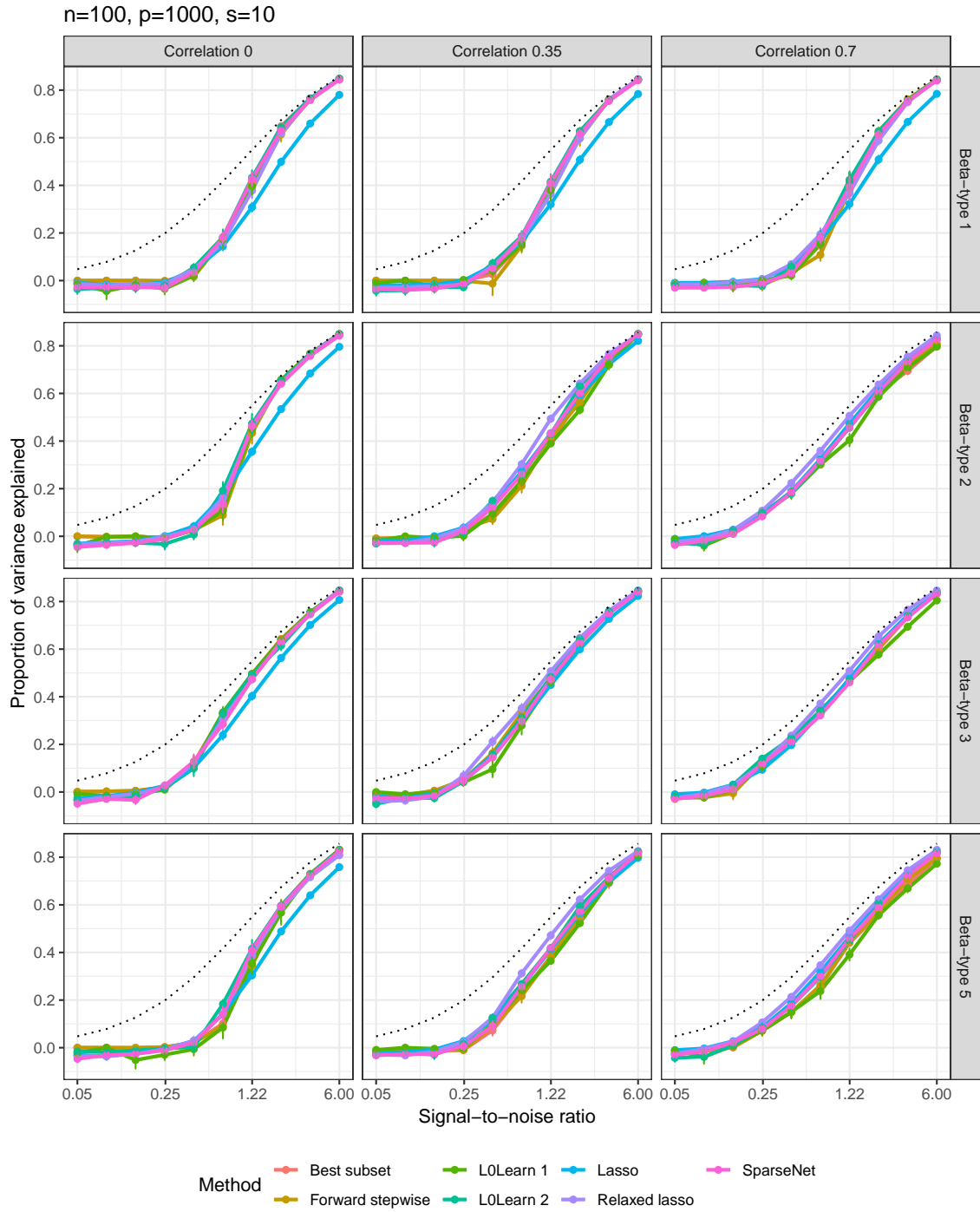


C.4 High-10 setting: $n = 100$, $p = 1000$, $s = 10$

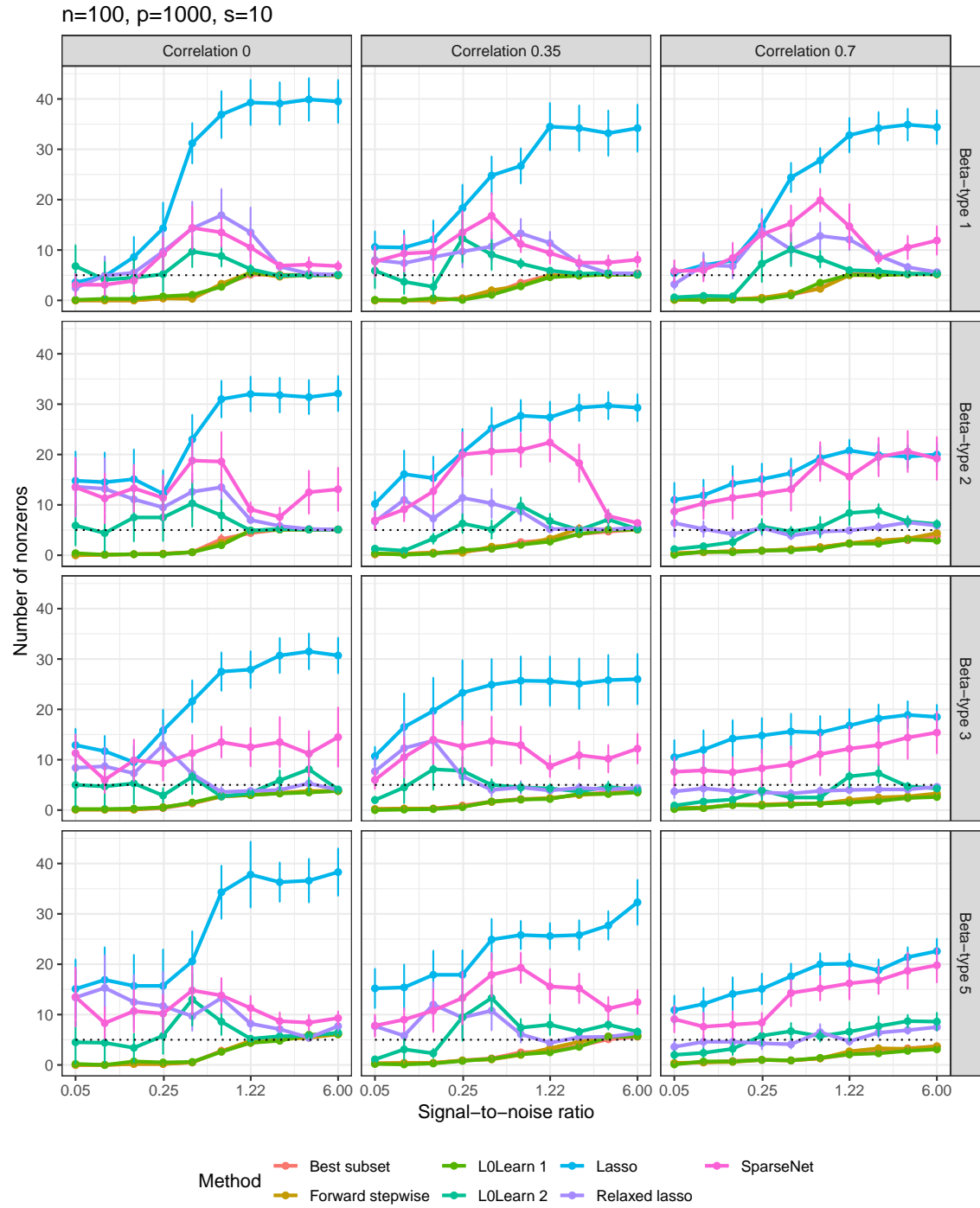
C.4.1 Relative test error (to Bayes)



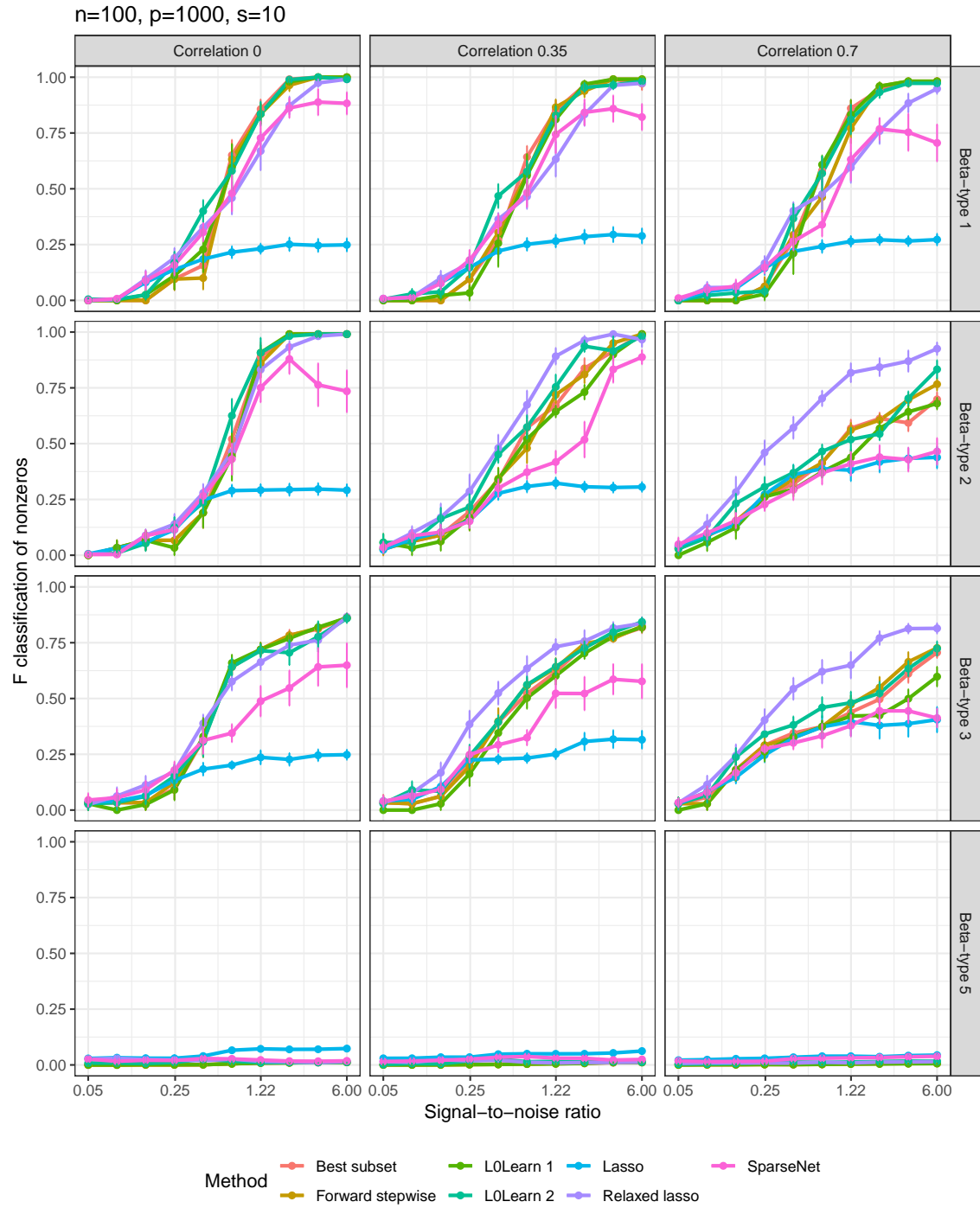
C.4.2 Proportion of variance explained



C.4.3 Number of nonzero coefficients



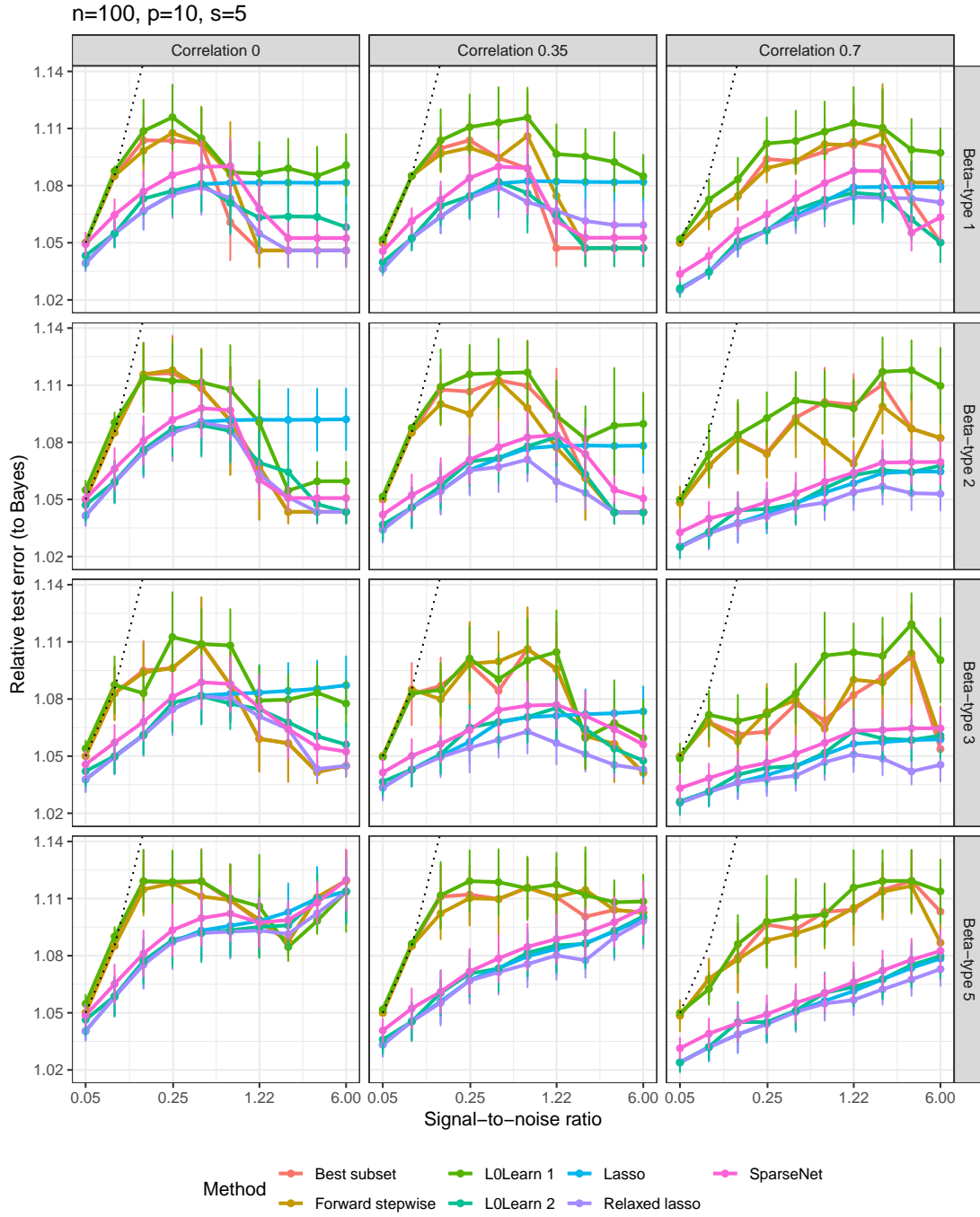
C.4.4 F-score



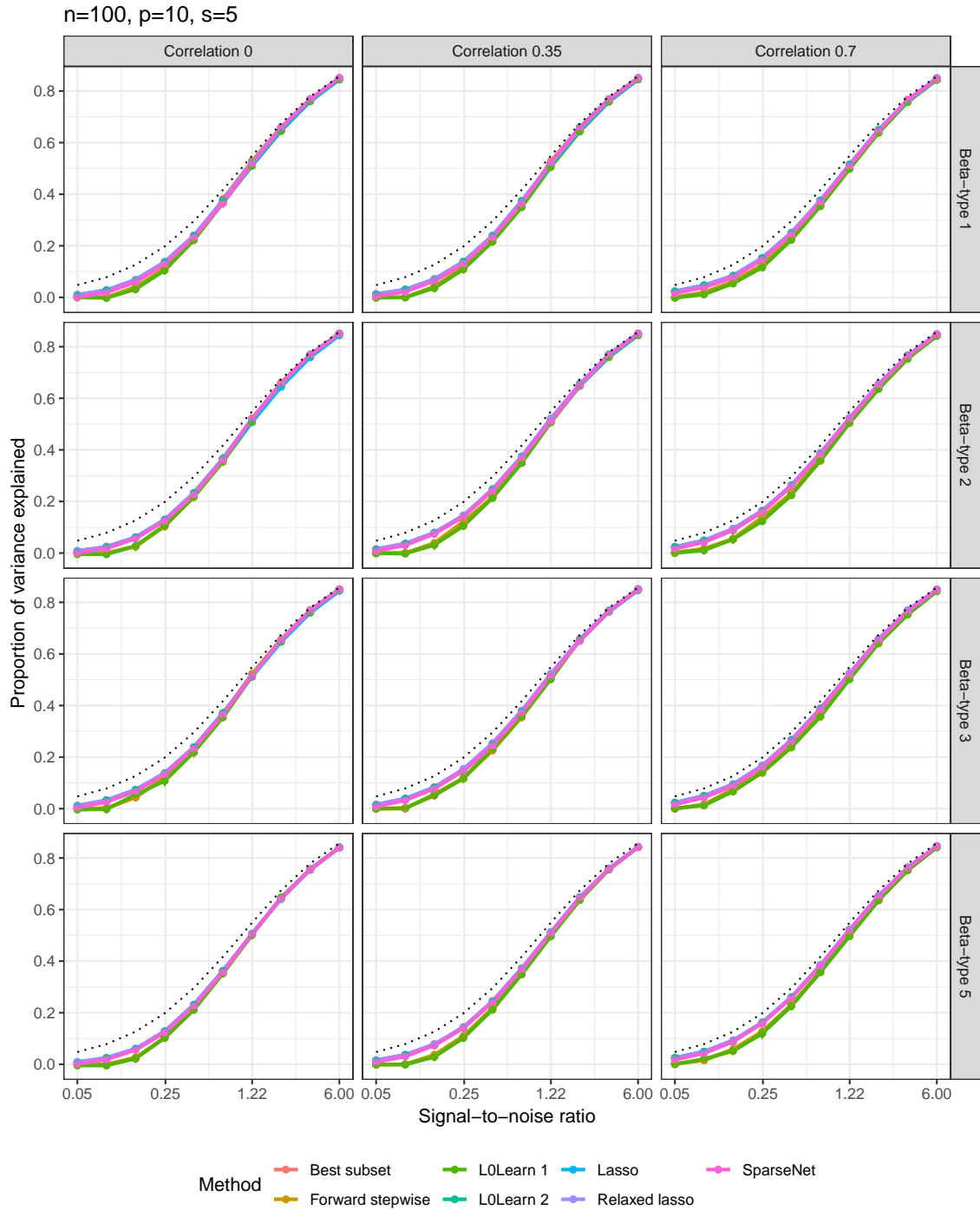
D Oracle tuning: now including L0Learn-1, L0Learn-2 and SparseNet

D.1 Low setting: $n = 100$, $p = 10$, $s = 5$

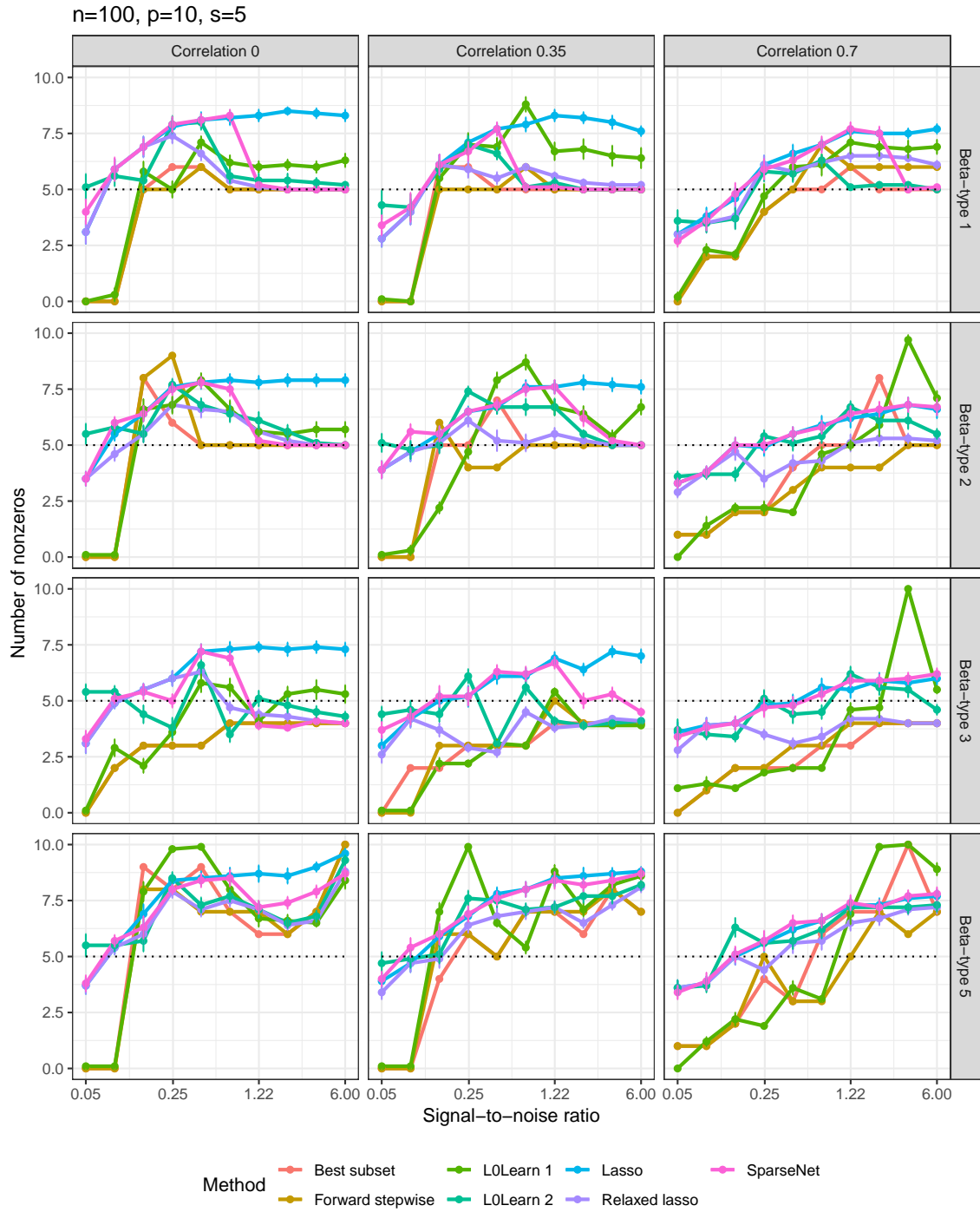
D.1.1 Relative test error (to Bayes)



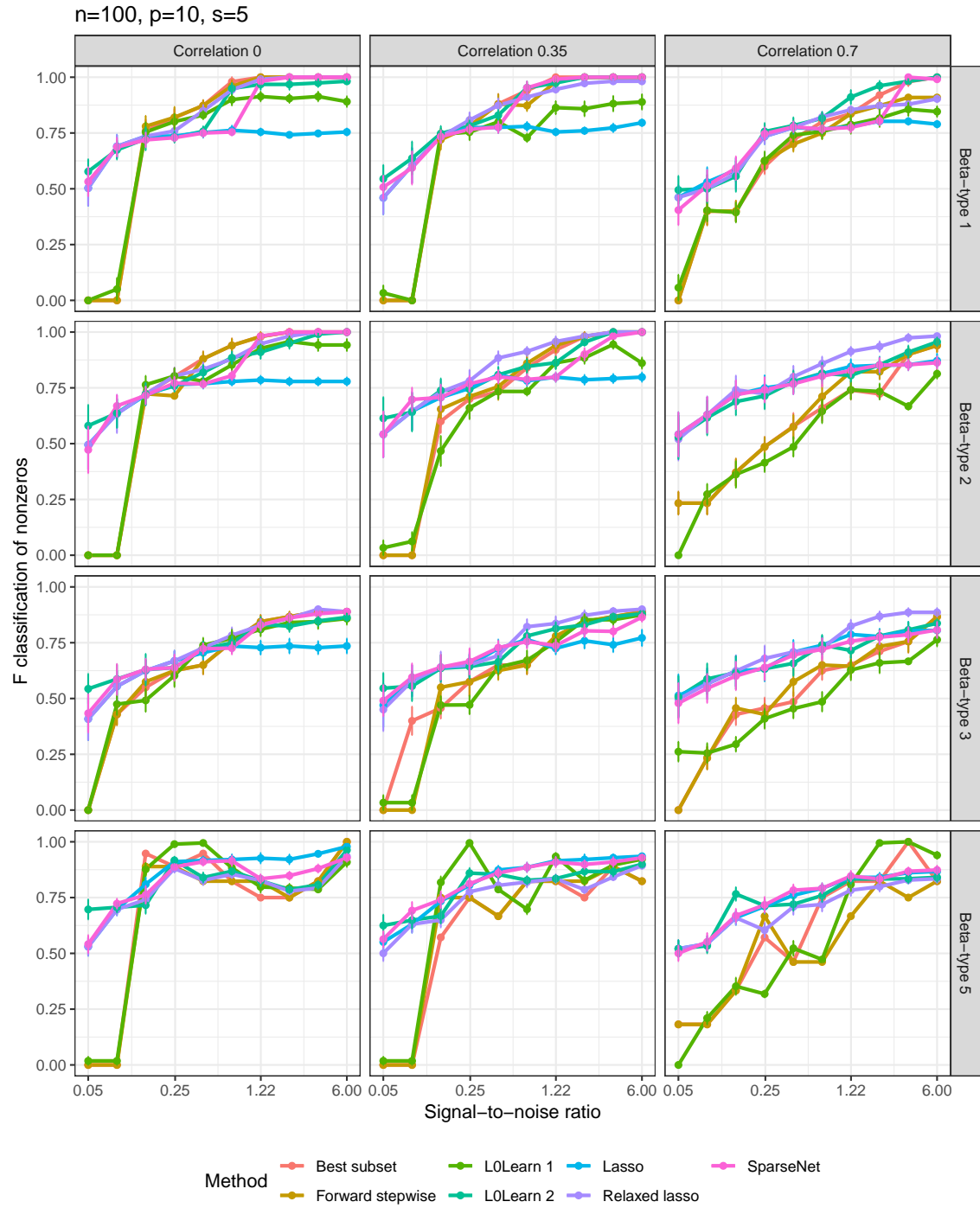
D.1.2 Proportion of variance explained



D.1.3 Number of nonzero coefficients

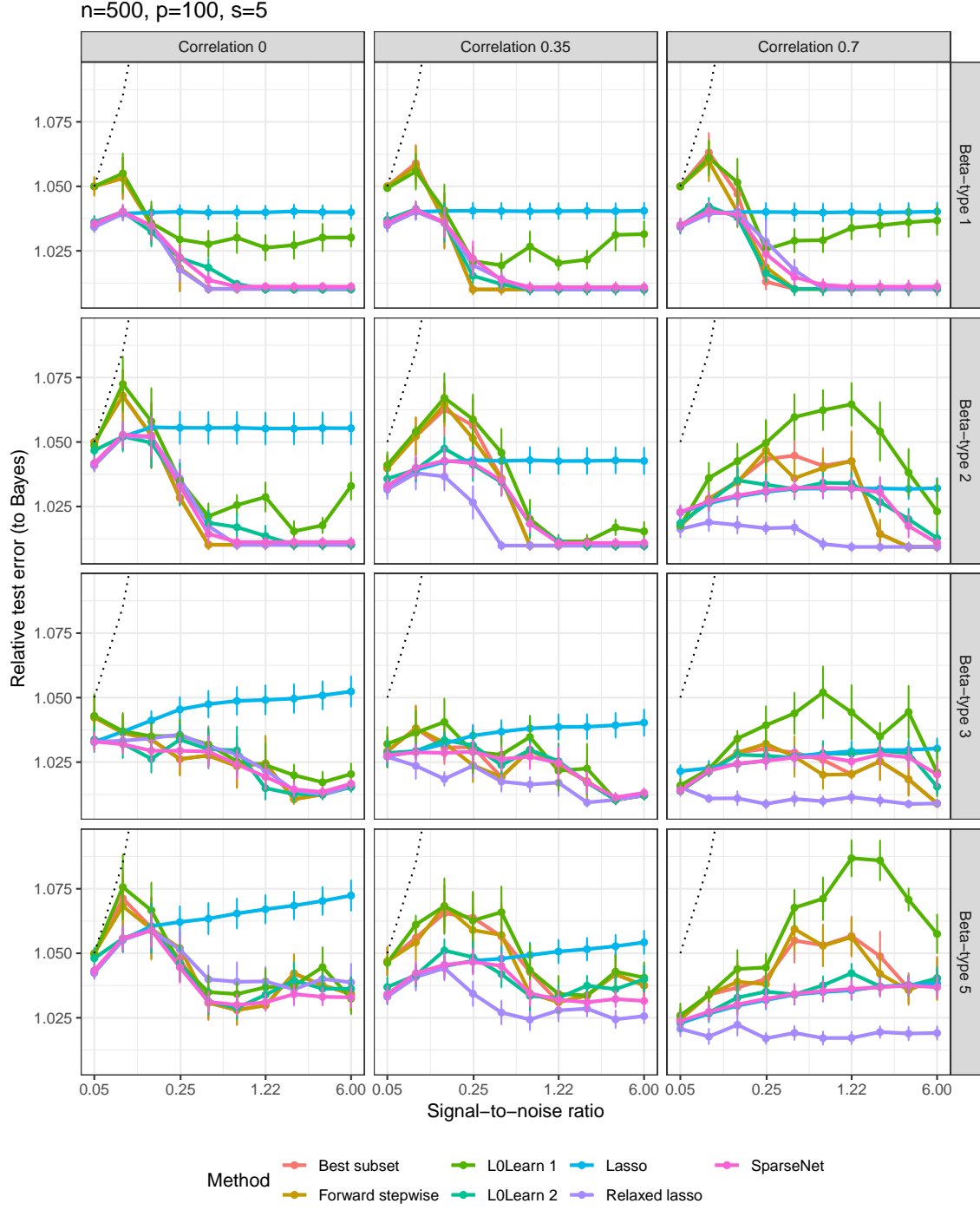


D.1.4 F-score

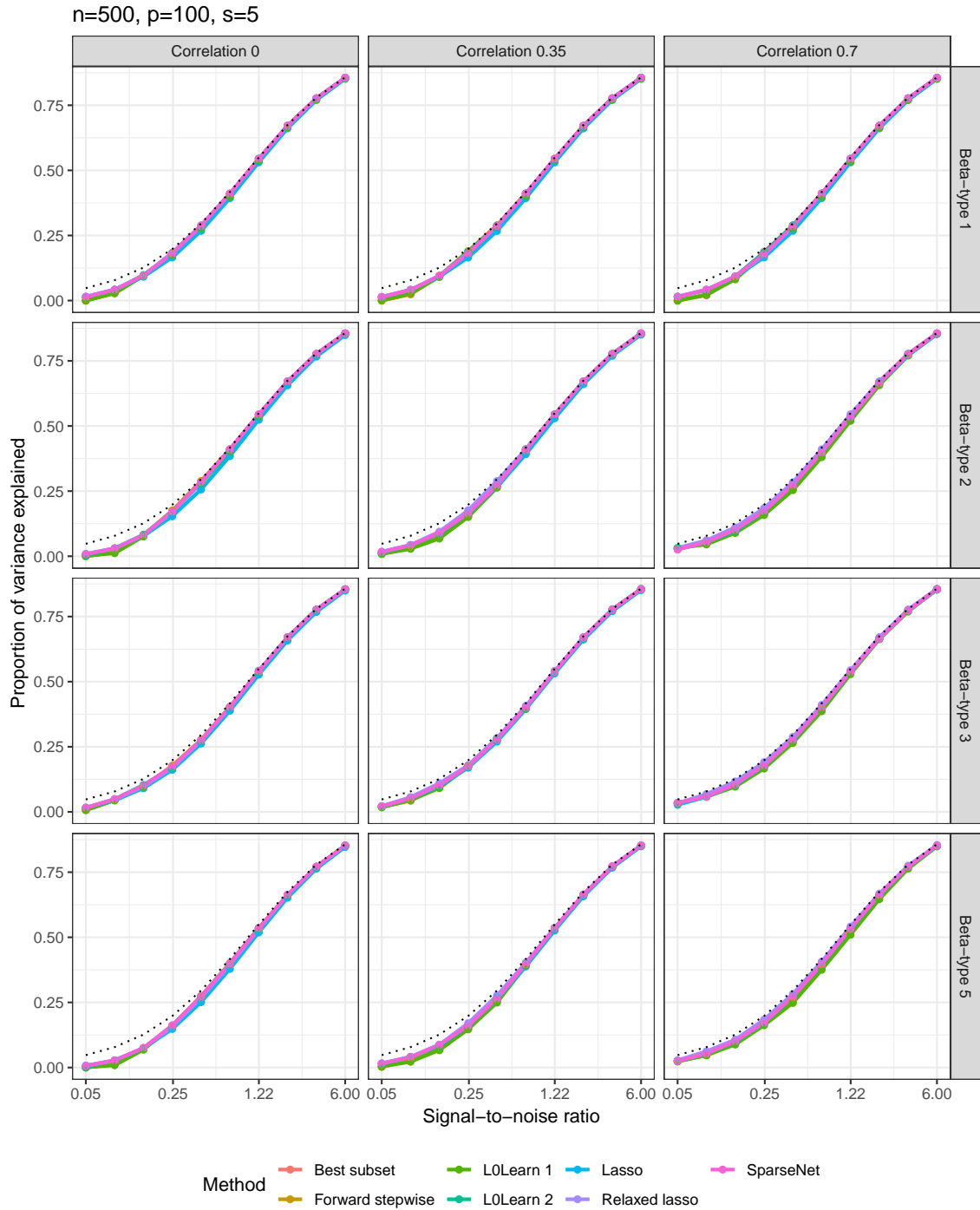


D.2 Medium setting: $n = 500$, $p = 100$, $s = 5$

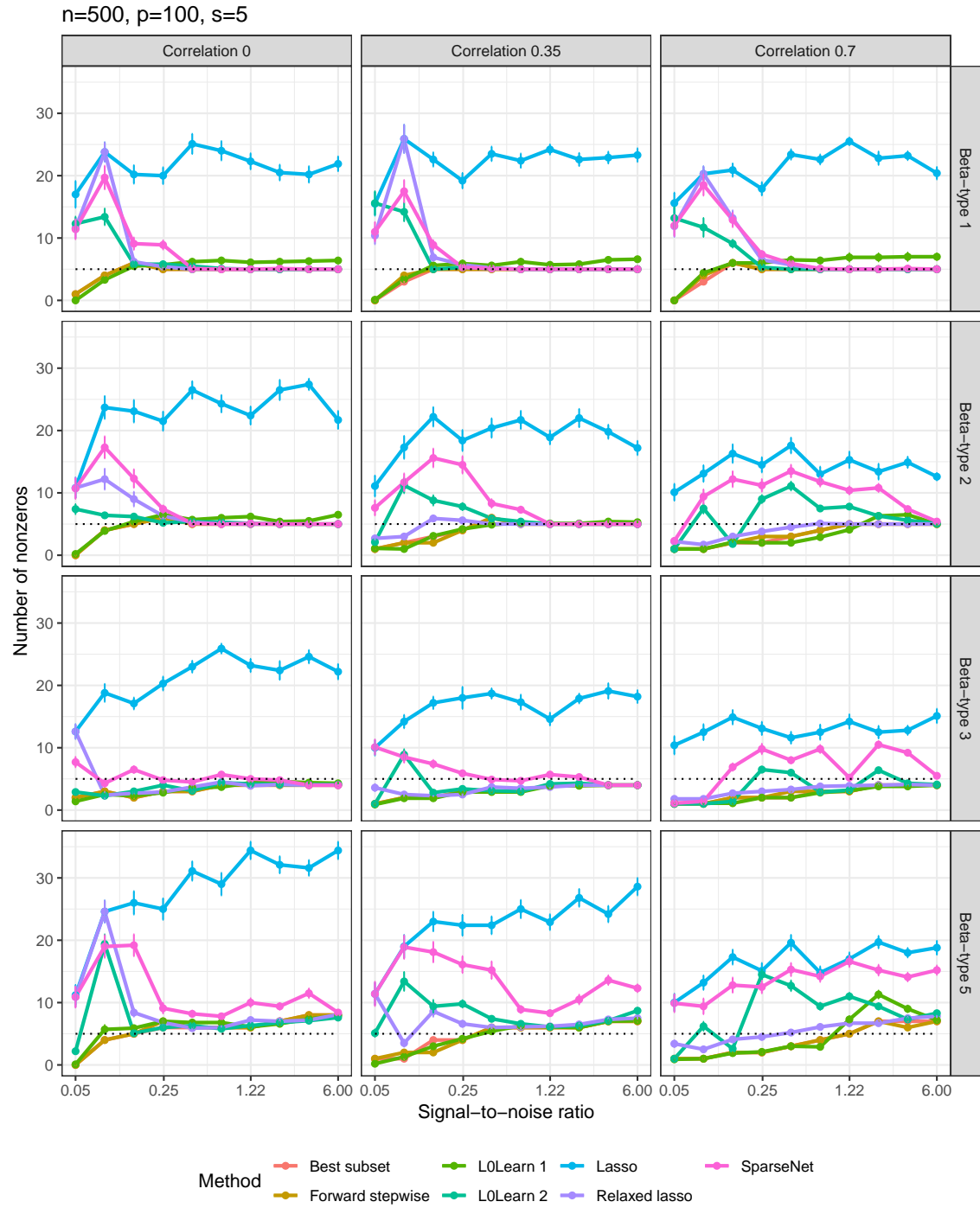
D.2.1 Relative test error (to Bayes)



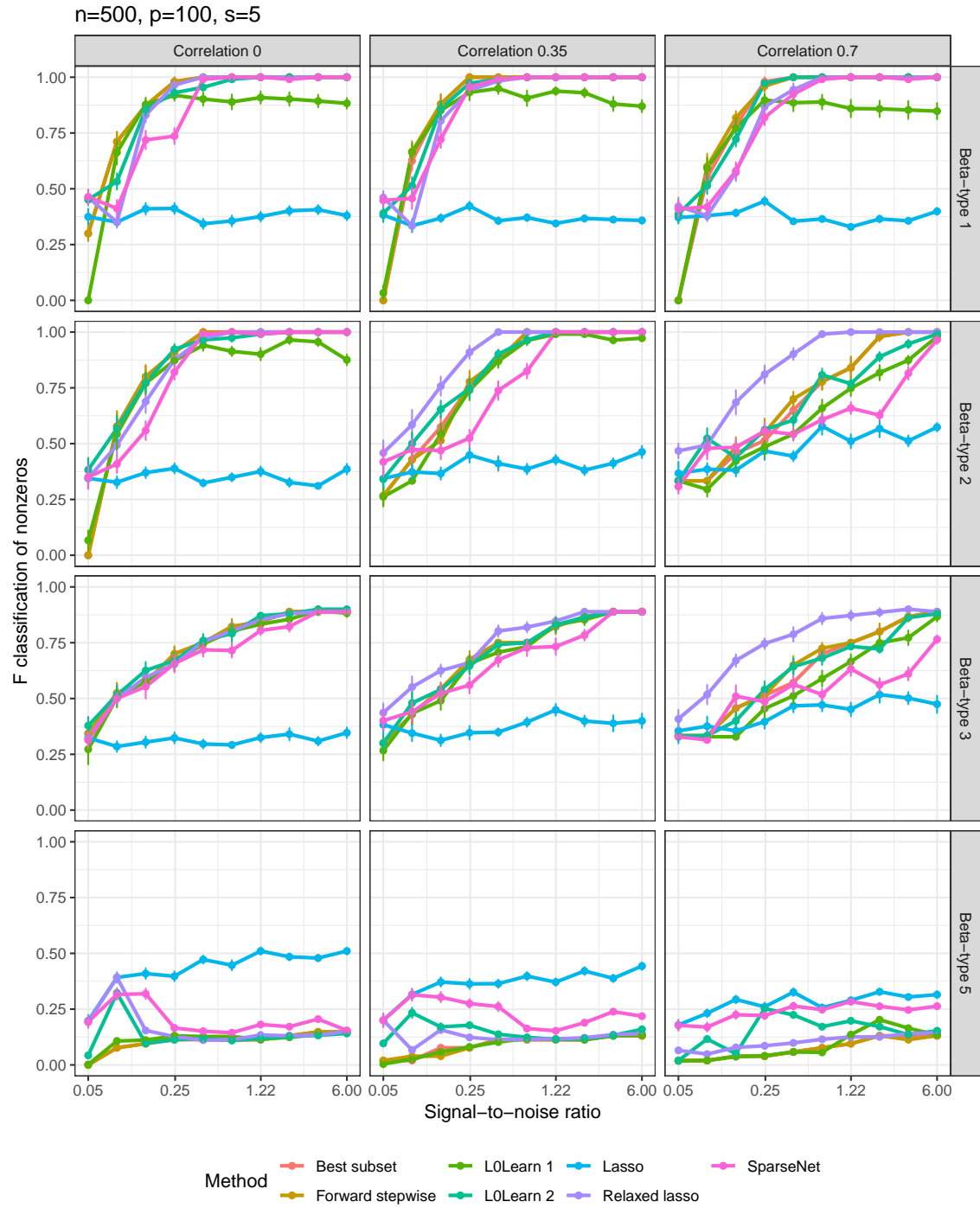
D.2.2 Proportion of variance explained



D.2.3 Number of nonzero coefficients

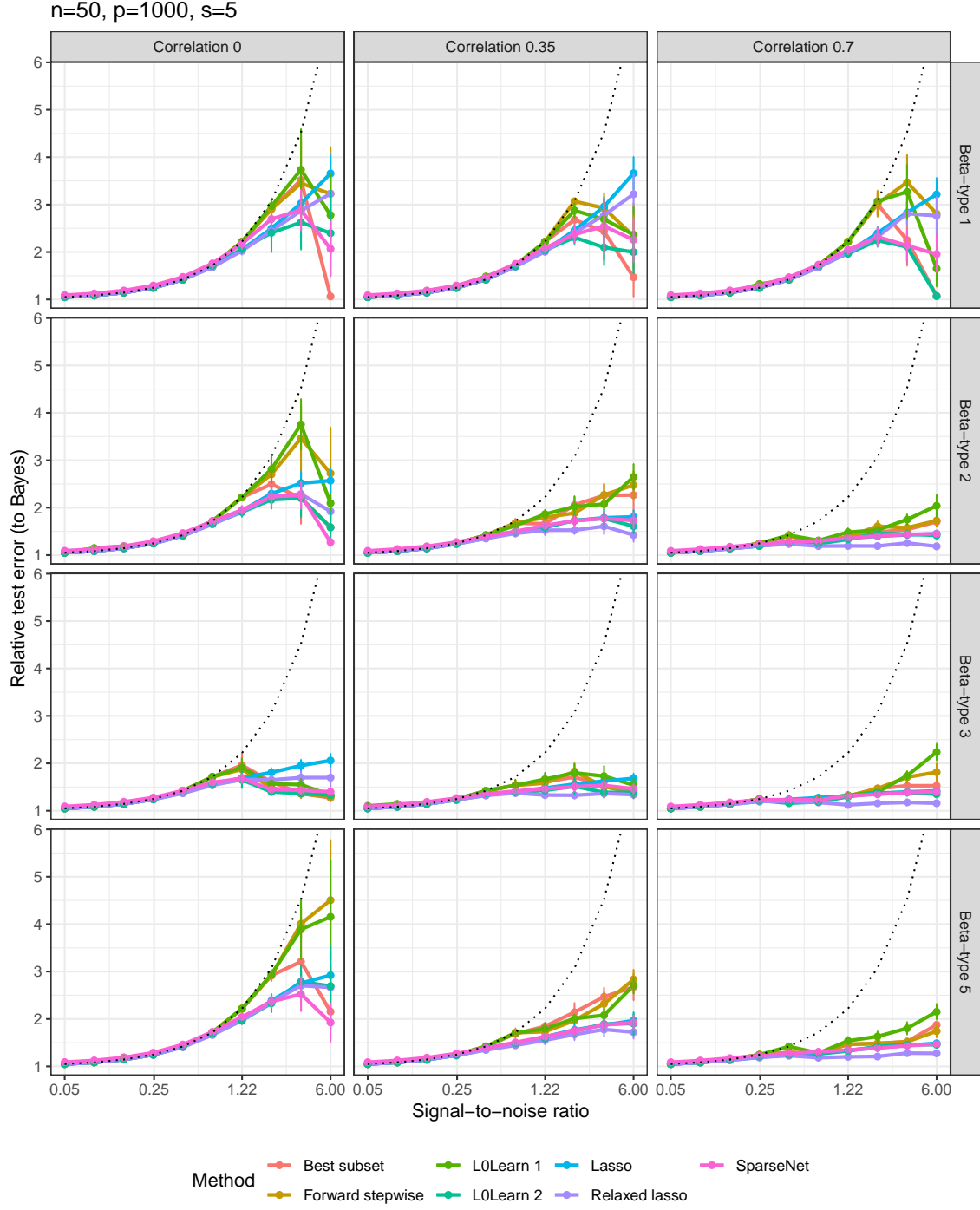


D.2.4 F-score

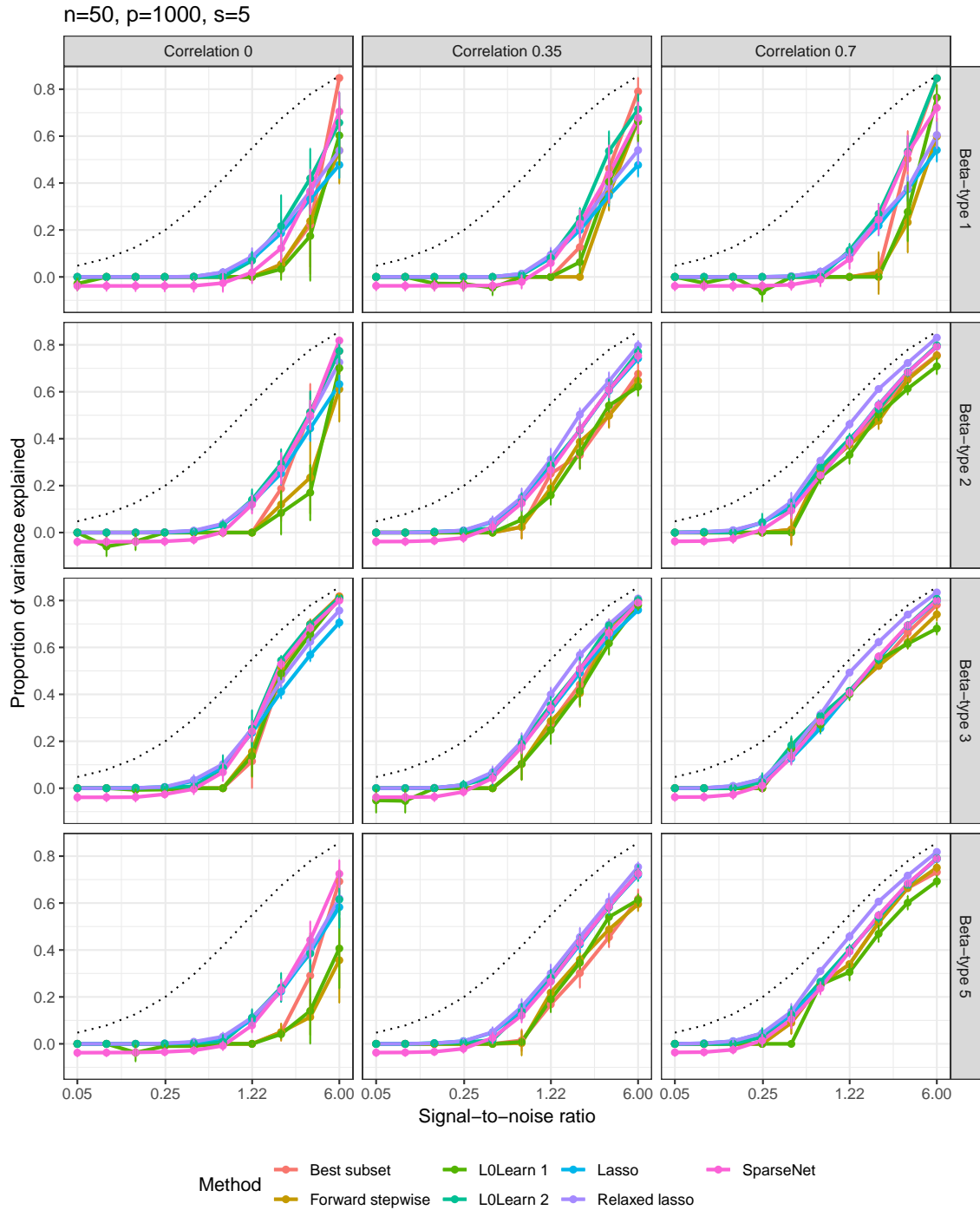


D.3 High-5 setting: $n = 50$, $p = 1000$, $s = 5$

D.3.1 Relative test error (to Bayes)



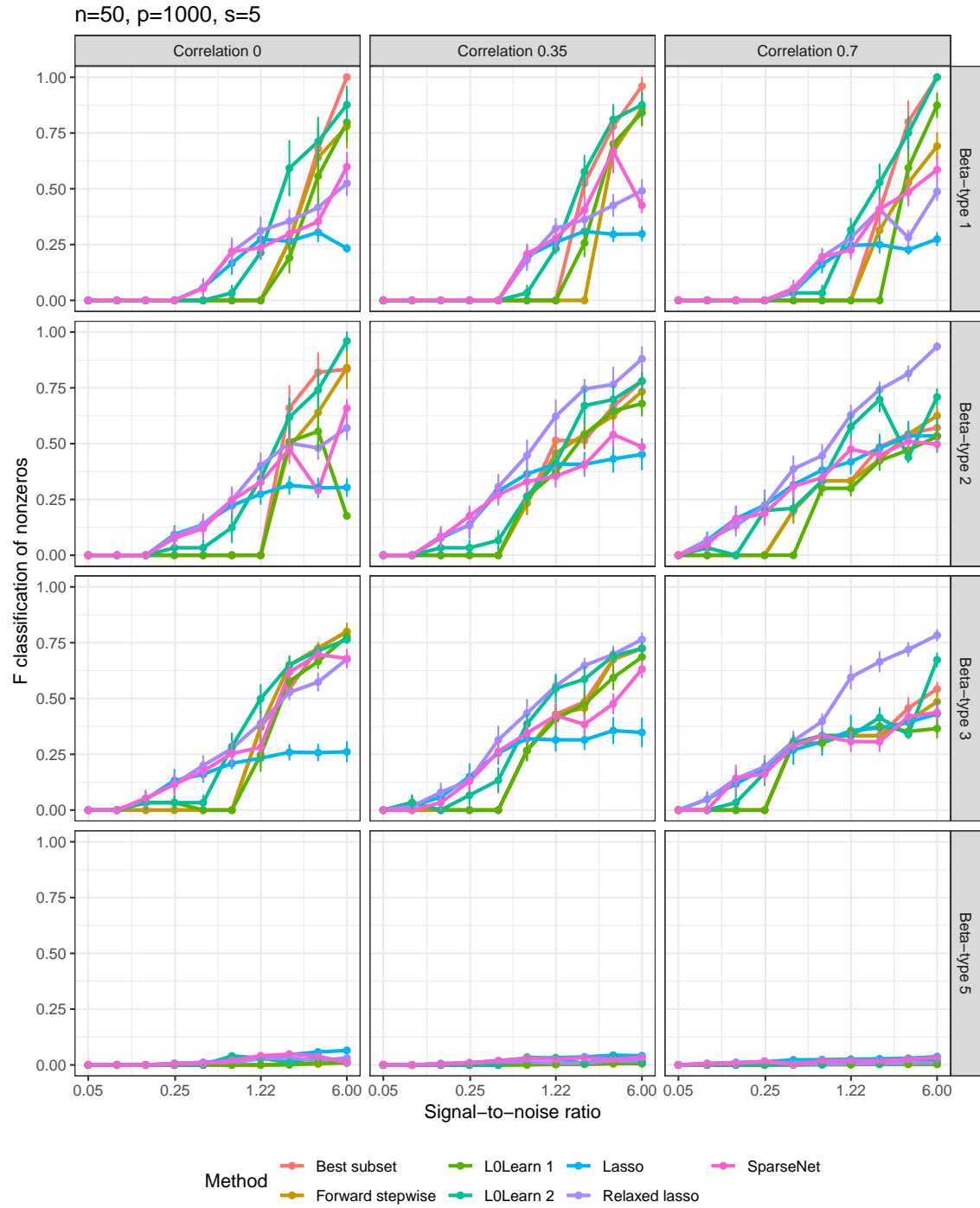
D.3.2 Proportion of variance explained



D.3.3 Number of nonzero coefficients



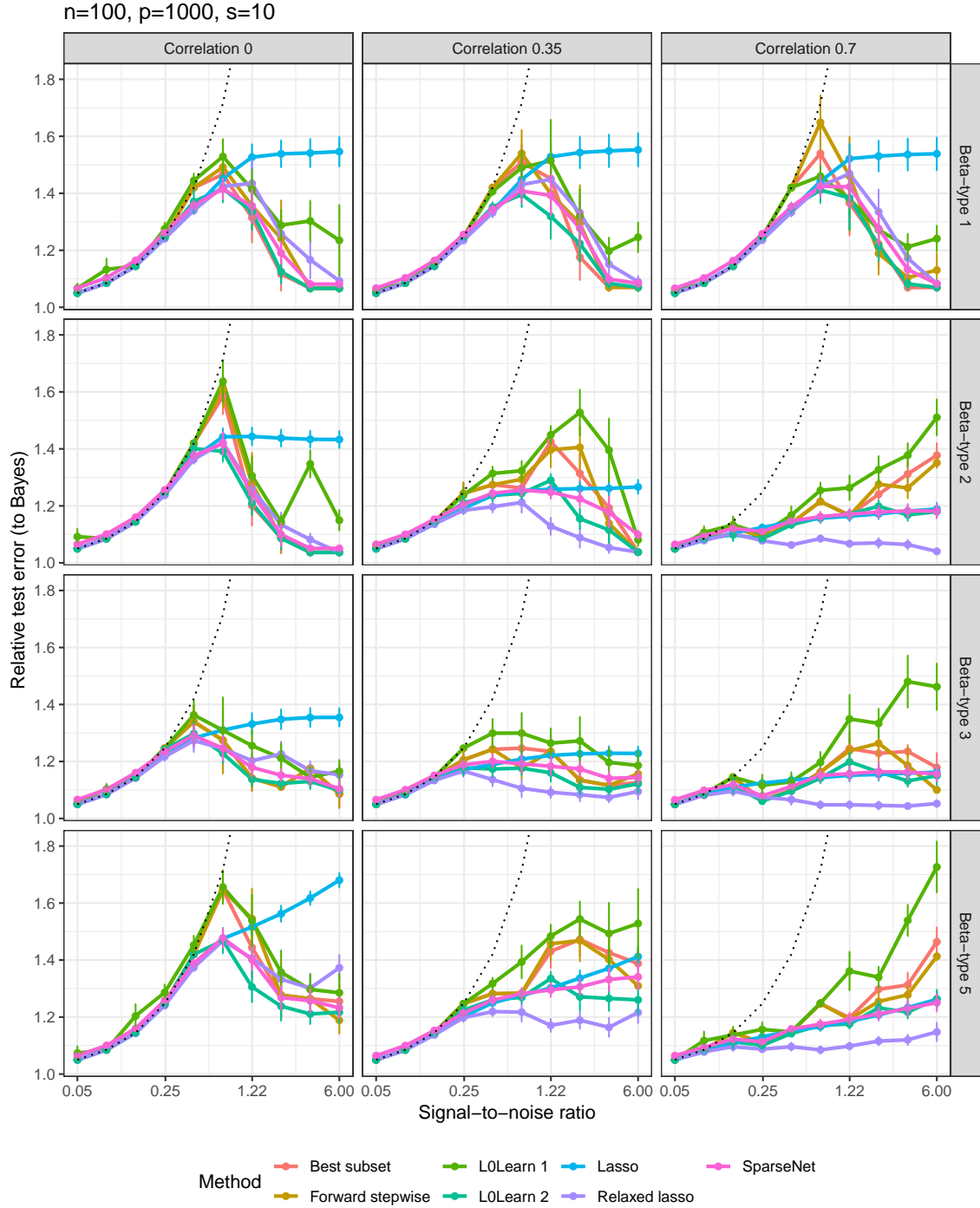
D.3.4 F-score



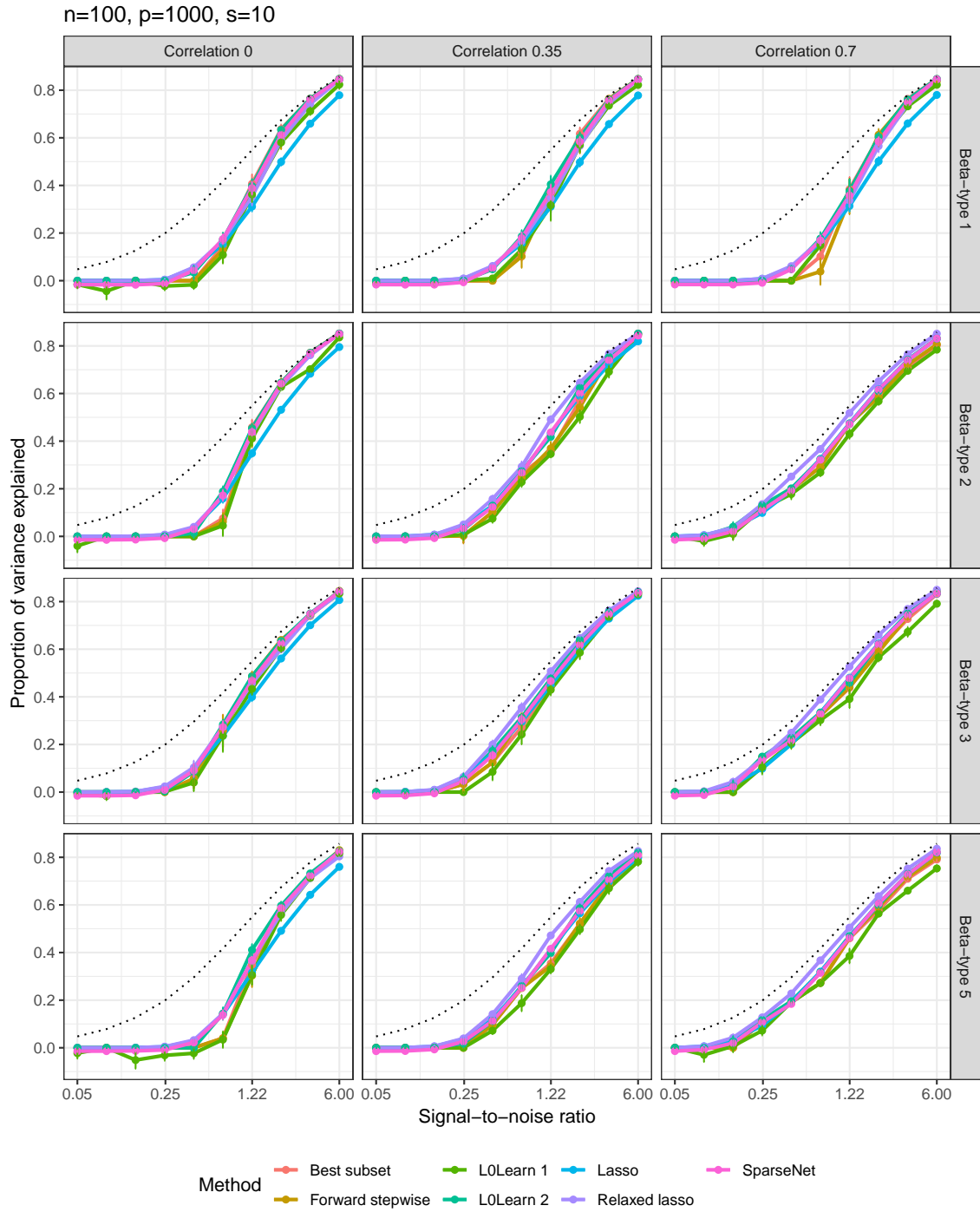
D.4 High-10 setting: $n = 100$, $p = 1000$, $s = 10$

D.4.1 Relative risk (to null model)

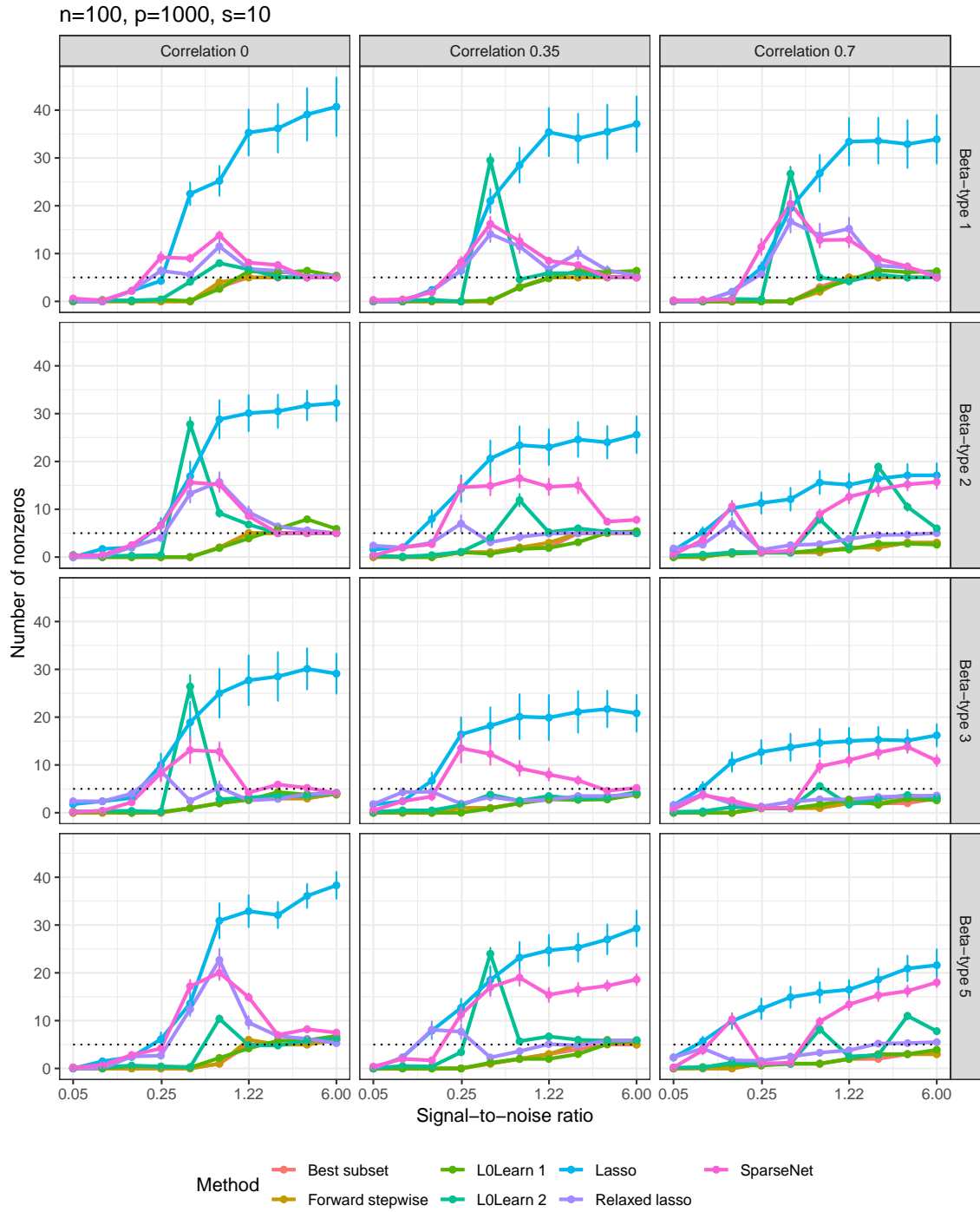
D.4.2 Relative test error (to Bayes)



D.4.3 Proportion of variance explained



D.4.4 Number of nonzero coefficients



D.4.5 F-score

