

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 (pure ℓ_0 penalty), L0Learn 2 (mixture of ℓ_0 and ℓ_1 penalties), and SparseNet.

Contents

A	Validation tuning	5
A.1	Low setting: $n = 100, p = 10, s = 5$	5
A.1.1	Relative risk (to null model)	5
A.1.2	Relative test error (to Bayes)	6
A.1.3	Proportion of variance explained	7
A.1.4	Number of nonzero coefficients	8
A.1.5	F-score	9
A.2	Medium setting: $n = 500, p = 100, s = 5$	10
A.2.1	Relative risk (to null model)	10
A.2.2	Relative test error (to Bayes)	11
A.2.3	Proportion of variance explained	12
A.2.4	Number of nonzero coefficients	13
A.2.5	F-score	14
A.3	High-5 setting: $n = 50, p = 1000, s = 5$	15
A.3.1	Relative risk (to null model)	15
A.3.2	Relative test error (to Bayes)	16
A.3.3	Proportion of variance explained	17
A.3.4	Number of nonzero coefficients	18
A.3.5	F-score	19
A.4	High-10 setting: $n = 100, p = 1000, s = 10$	20
A.4.1	Relative risk (to null model)	20
A.4.2	Relative test error (to Bayes)	21
A.4.3	Proportion of variance explained	22
A.4.4	Number of nonzero coefficients	23
A.4.5	F-score	24
B	Oracle tuning	25
B.1	Low setting: $n = 100, p = 10, s = 5$	25
B.1.1	Relative risk (to null model)	25
B.1.2	Relative test error (to Bayes)	26
B.1.3	Proportion of variance explained	27
B.1.4	Number of nonzero coefficients	28
B.1.5	F-score	29
B.2	Medium setting: $n = 500, p = 100, s = 5$	30
B.2.1	Relative risk (to null model)	30
B.2.2	Relative test error (to Bayes)	31
B.2.3	Proportion of variance explained	32
B.2.4	Number of nonzero coefficients	33
B.2.5	F-score	34
B.3	High-5 setting: $n = 50, p = 1000, s = 5$	35
B.3.1	Relative risk (to null model)	35
B.3.2	Relative test error (to Bayes)	36
B.3.3	Proportion of variance explained	37
B.3.4	Number of nonzero coefficients	38
B.3.5	F-score	39
B.4	High-10 setting: $n = 100, p = 1000, s = 10$	40
B.4.1	Relative risk (to null model)	40
B.4.2	Relative test error (to Bayes)	41
B.4.3	Proportion of variance explained	42

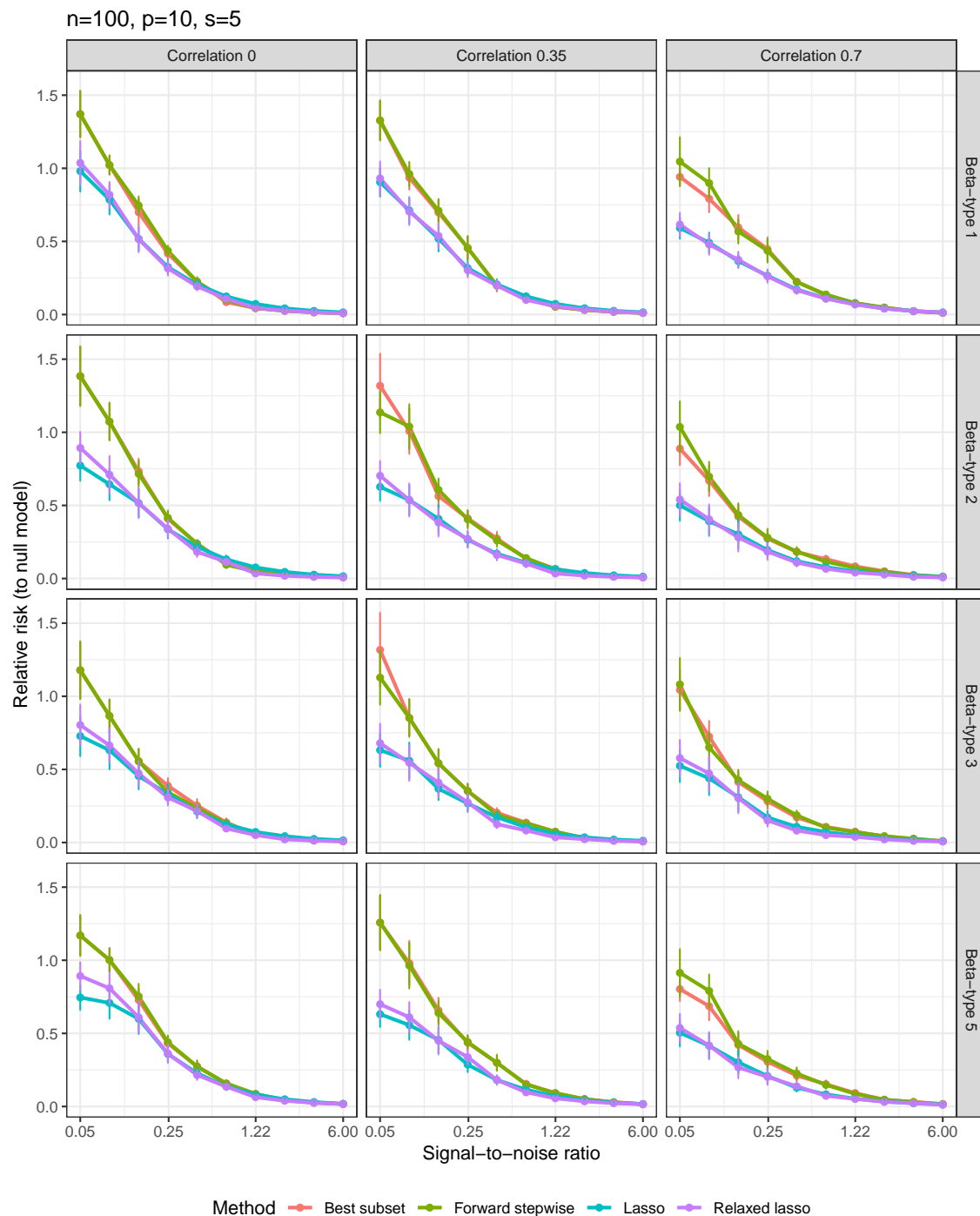
B.4.4	Number of nonzero coefficients	43
B.4.5	F-score	44
C	Validation tuning: more methods	45
C.1	Low setting: $n = 100, p = 10, s = 5$	45
C.1.1	Relative risk (to null model)	45
C.1.2	Relative test error (to Bayes)	46
C.1.3	Proportion of variance explained	47
C.1.4	Number of nonzero coefficients	48
C.1.5	F-score	49
C.2	Medium setting: $n = 500, p = 100, s = 5$	50
C.2.1	Relative risk (to null model)	50
C.2.2	Relative test error (to Bayes)	51
C.2.3	Proportion of variance explained	52
C.2.4	Number of nonzero coefficients	53
C.2.5	F-score	54
C.3	High-5 setting: $n = 50, p = 1000, s = 5$	55
C.3.1	Relative risk (to null model)	55
C.3.2	Relative test error (to Bayes)	56
C.3.3	Proportion of variance explained	57
C.3.4	Number of nonzero coefficients	58
C.3.5	F-score	59
C.4	High-10 setting: $n = 100, p = 1000, s = 10$	60
C.4.1	Relative risk (to null model)	60
C.4.2	Relative test error (to Bayes)	61
C.4.3	Proportion of variance explained	62
C.4.4	Number of nonzero coefficients	63
C.4.5	F-score	64
D	Oracle tuning: more methods	65
D.1	Low setting: $n = 100, p = 10, s = 5$	65
D.1.1	Relative risk (to null model)	65
D.1.2	Relative test error (to Bayes)	66
D.1.3	Proportion of variance explained	67
D.1.4	Number of nonzero coefficients	68
D.1.5	F-score	69
D.2	Medium setting: $n = 500, p = 100, s = 5$	70
D.2.1	Relative risk (to null model)	70
D.2.2	Relative test error (to Bayes)	71
D.2.3	Proportion of variance explained	72
D.2.4	Number of nonzero coefficients	73
D.2.5	F-score	74
D.3	High-5 setting: $n = 50, p = 1000, s = 5$	75
D.3.1	Relative risk (to null model)	75
D.3.2	Relative test error (to Bayes)	76
D.3.3	Proportion of variance explained	77
D.3.4	Number of nonzero coefficients	78
D.3.5	F-score	79
D.4	High-10 setting: $n = 100, p = 1000, s = 10$	80
D.4.1	Relative risk (to null model)	80
D.4.2	Relative test error (to Bayes)	81

D.4.3	Proportion of variance explained	82
D.4.4	Number of nonzero coefficients	83
D.4.5	F-score	84

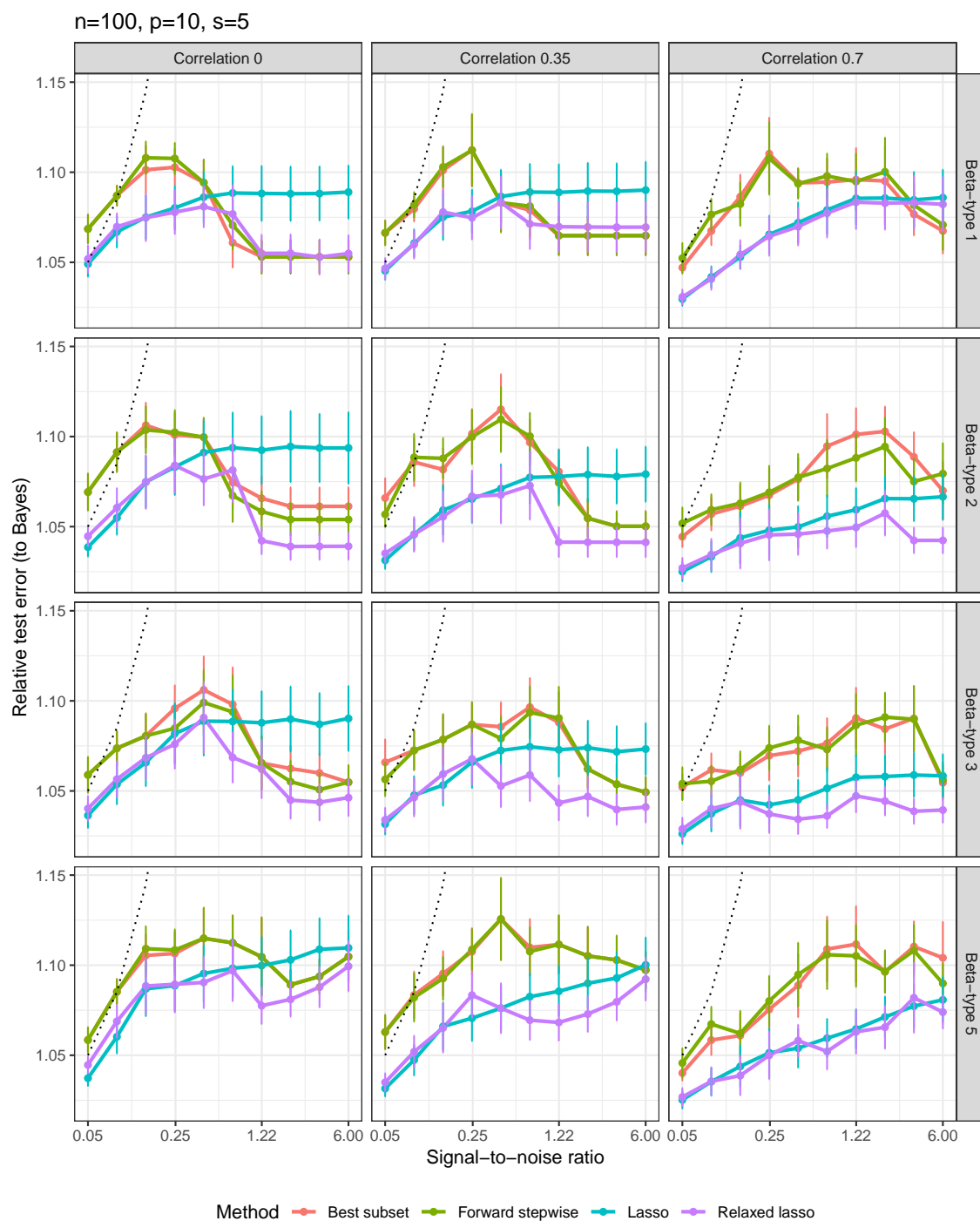
A Validation tuning

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

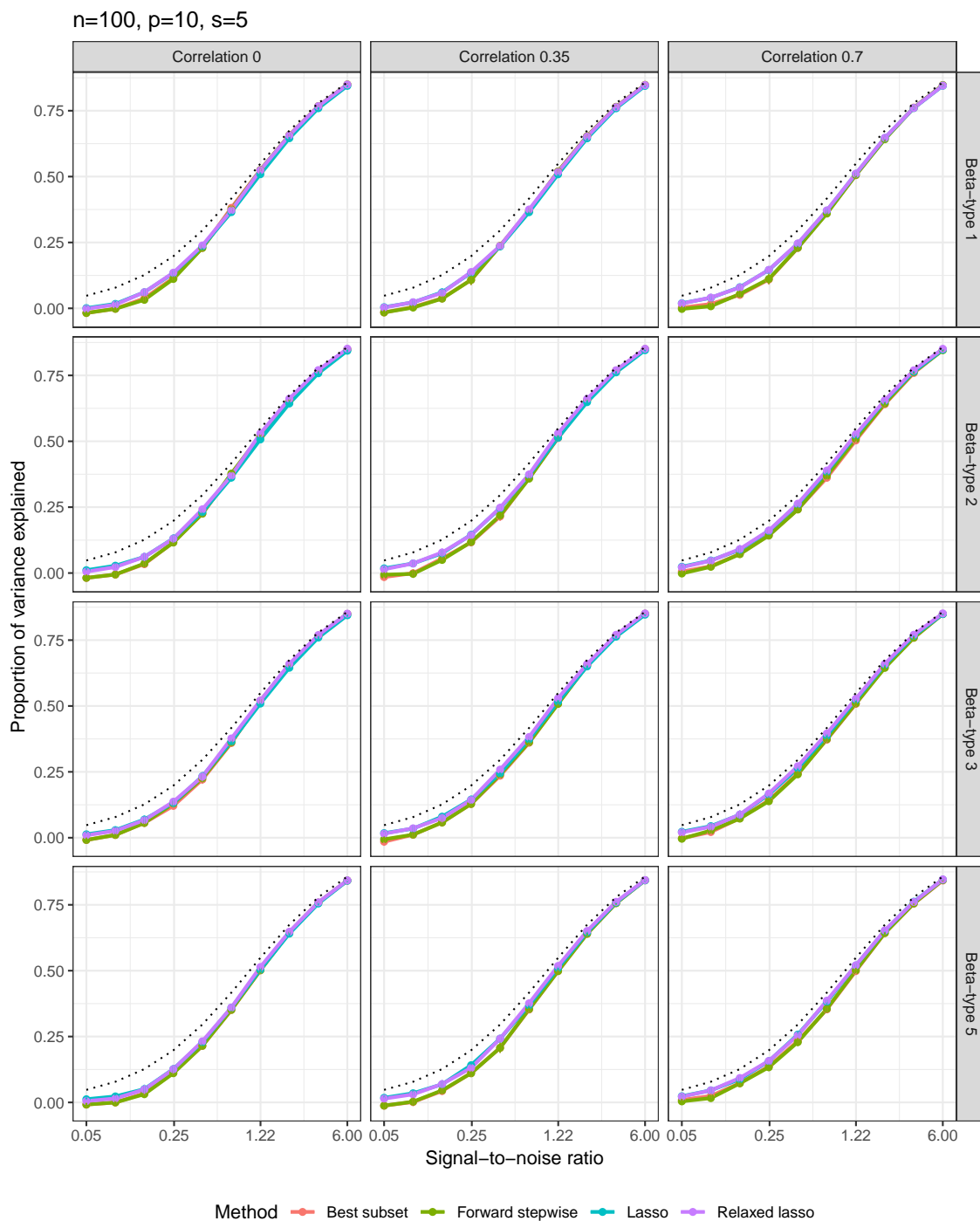
A.1.1 Relative risk (to null model)



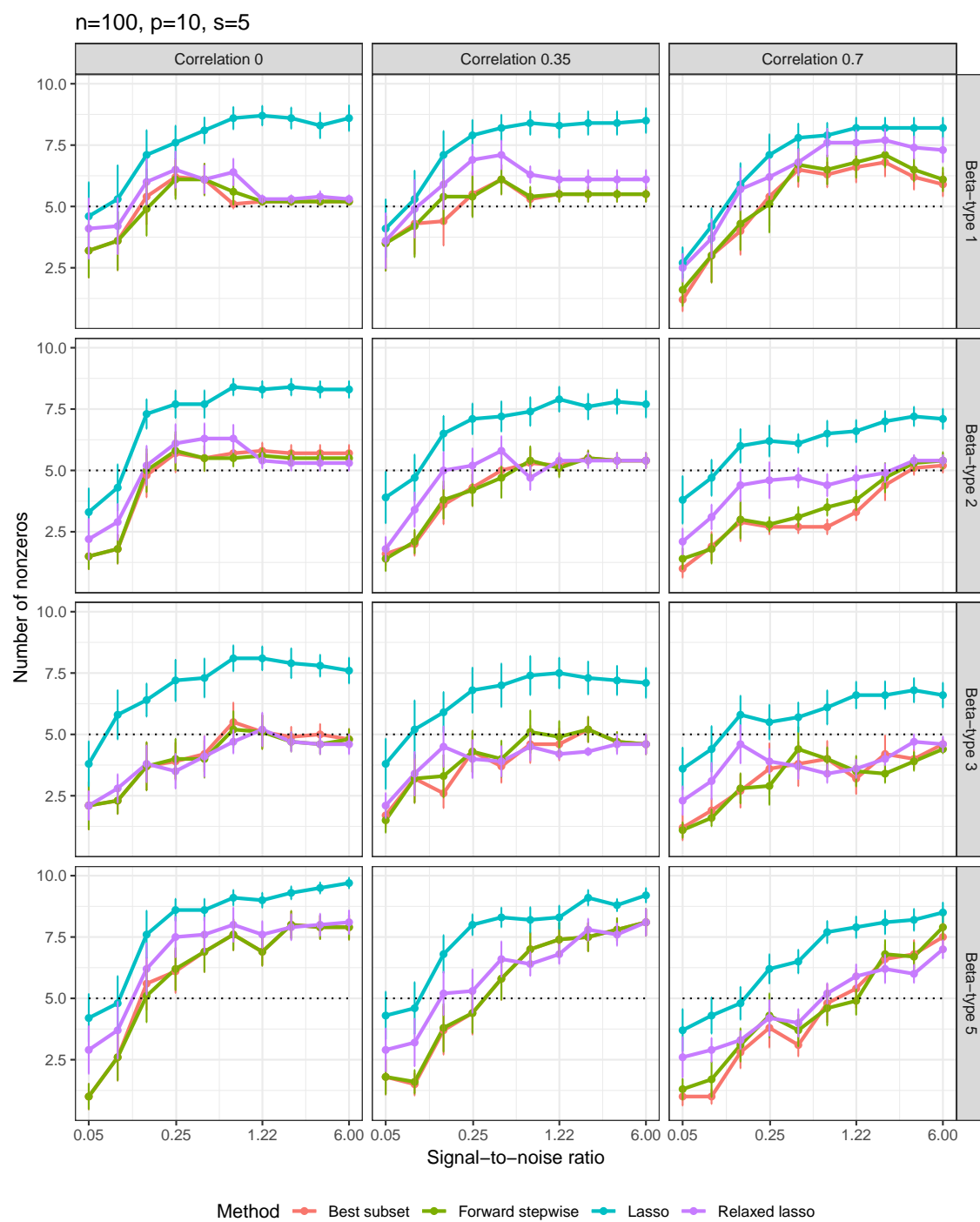
A.1.2 Relative test error (to Bayes)



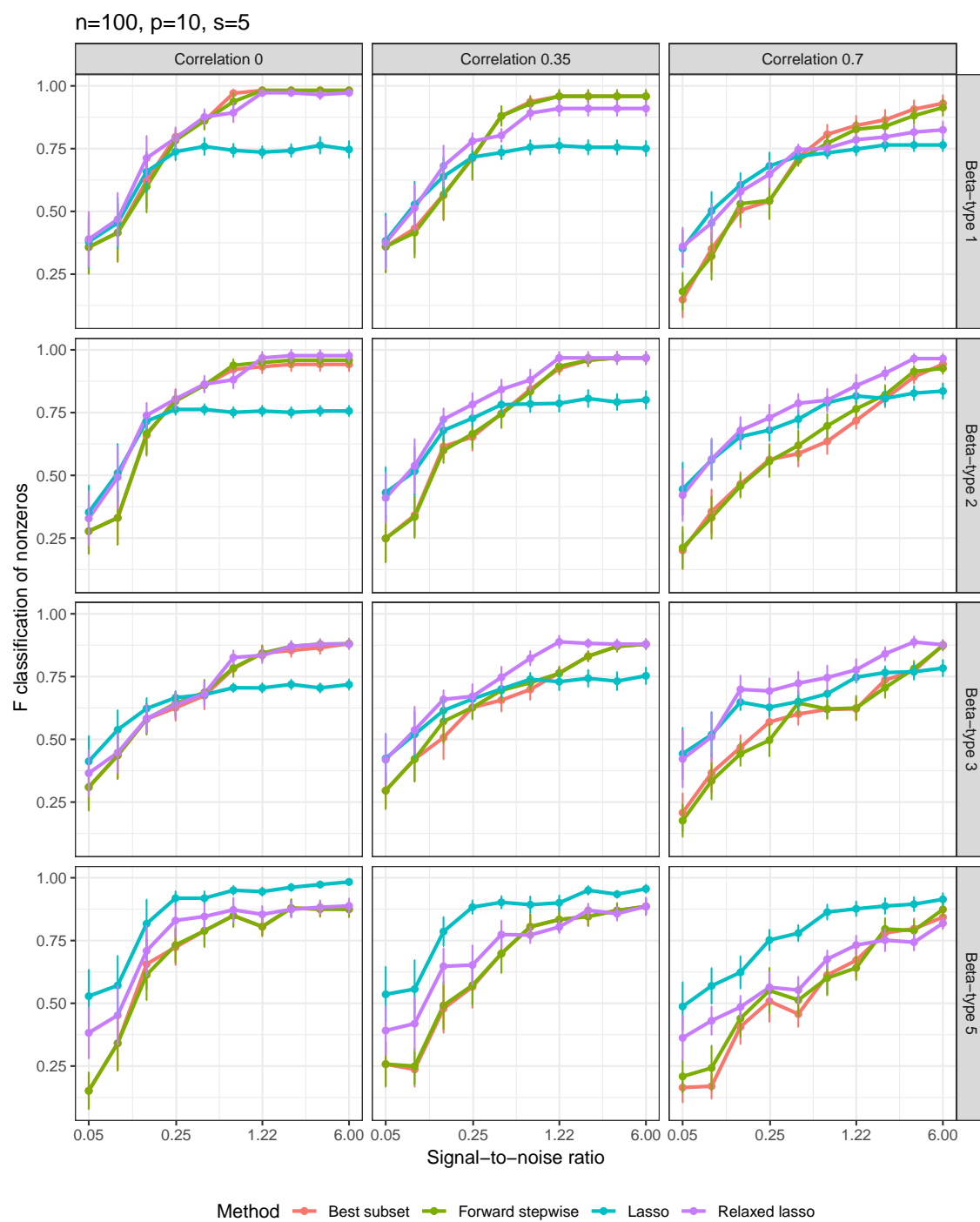
A.1.3 Proportion of variance explained



A.1.4 Number of nonzero coefficients

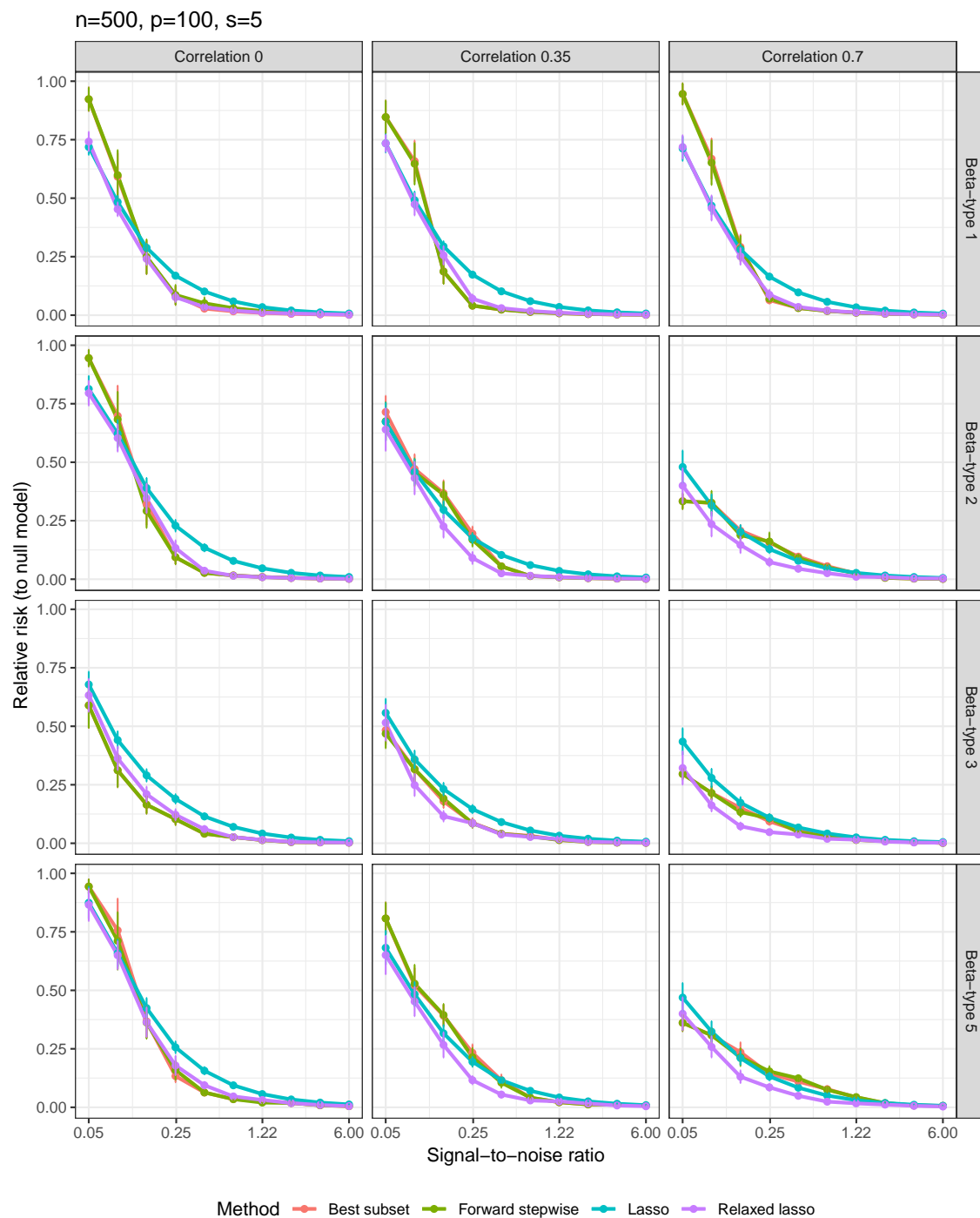


A.1.5 F-score

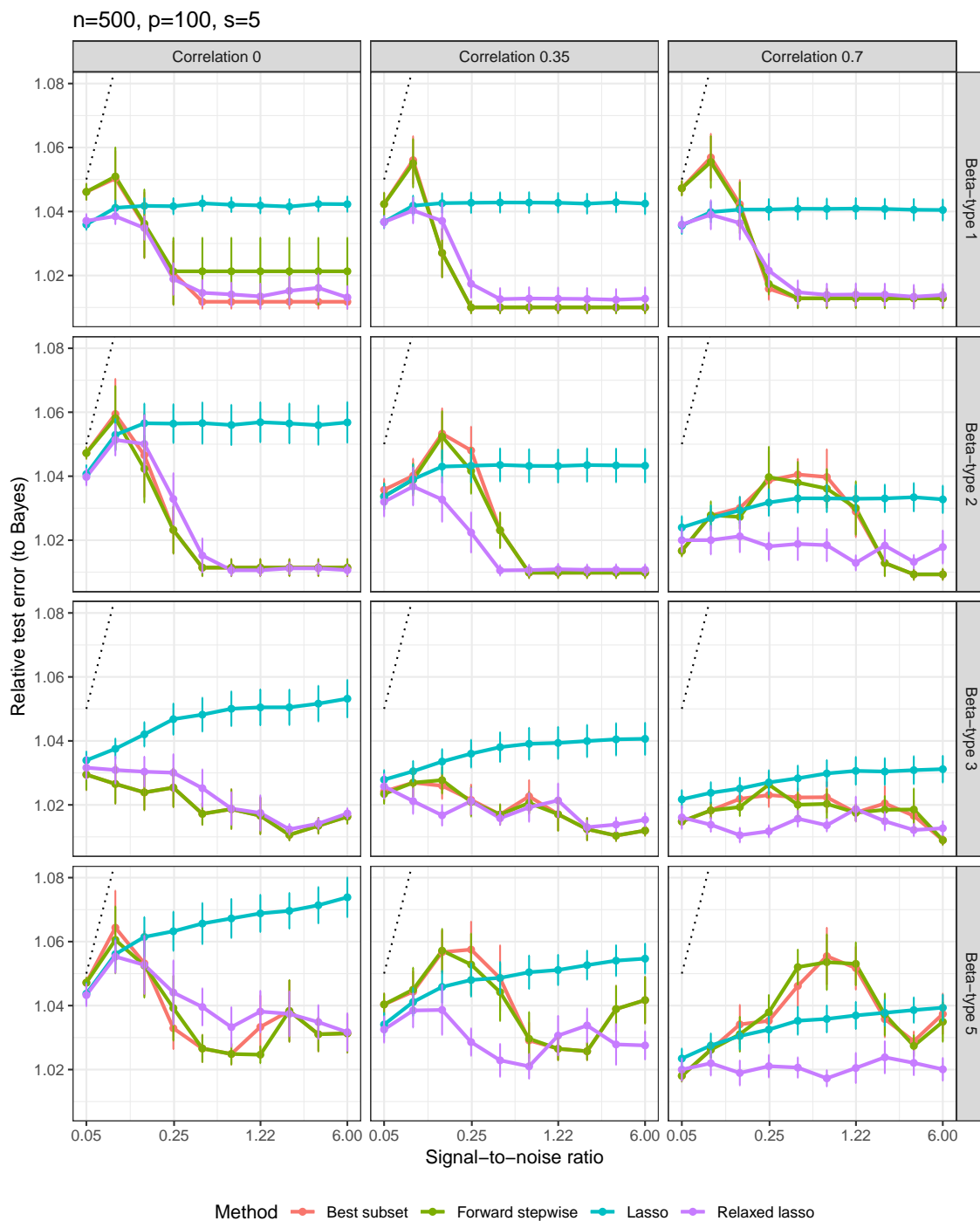


A.2 Medium setting: $n = 500, p = 100, s = 5$

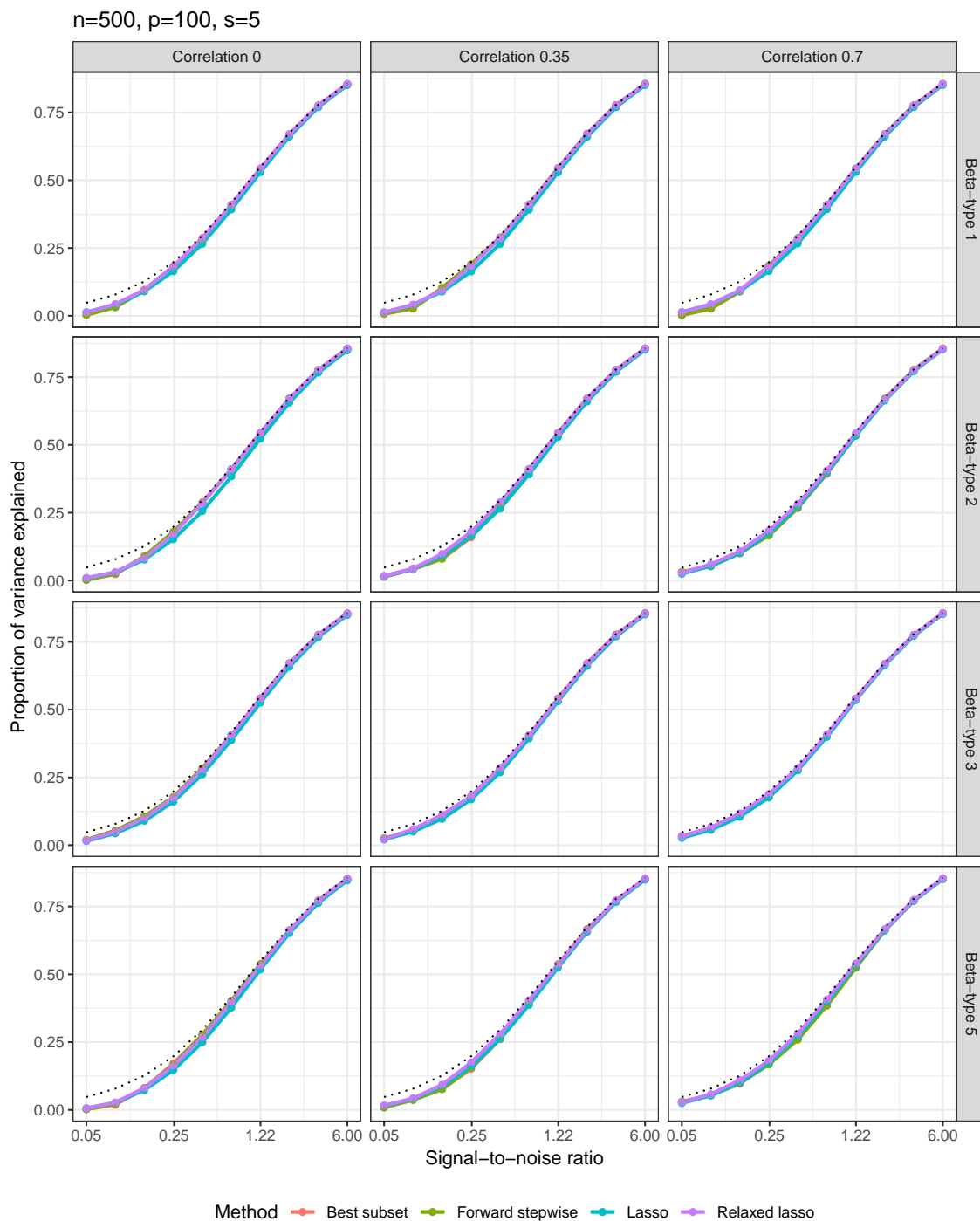
A.2.1 Relative risk (to null model)



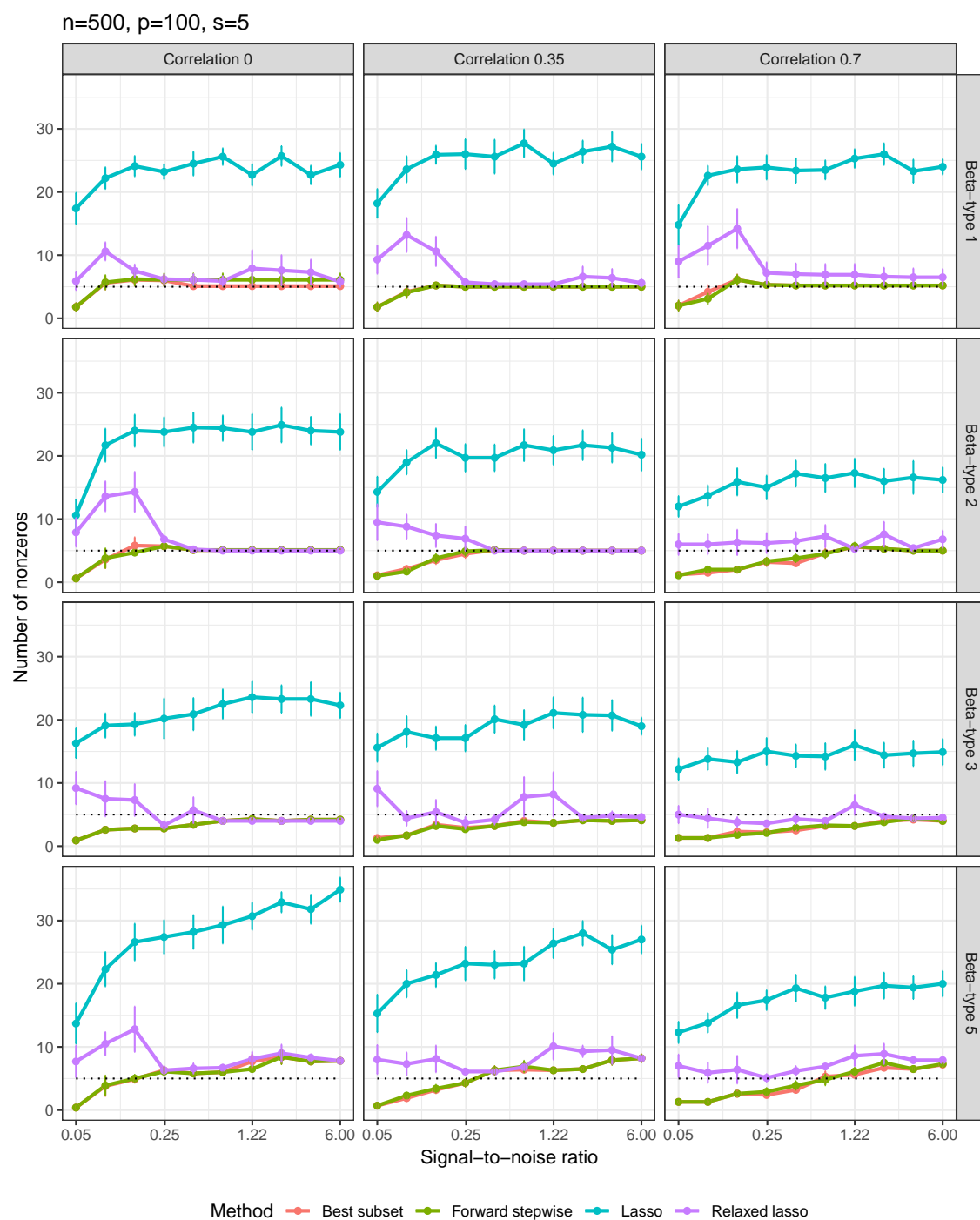
A.2.2 Relative test error (to Bayes)



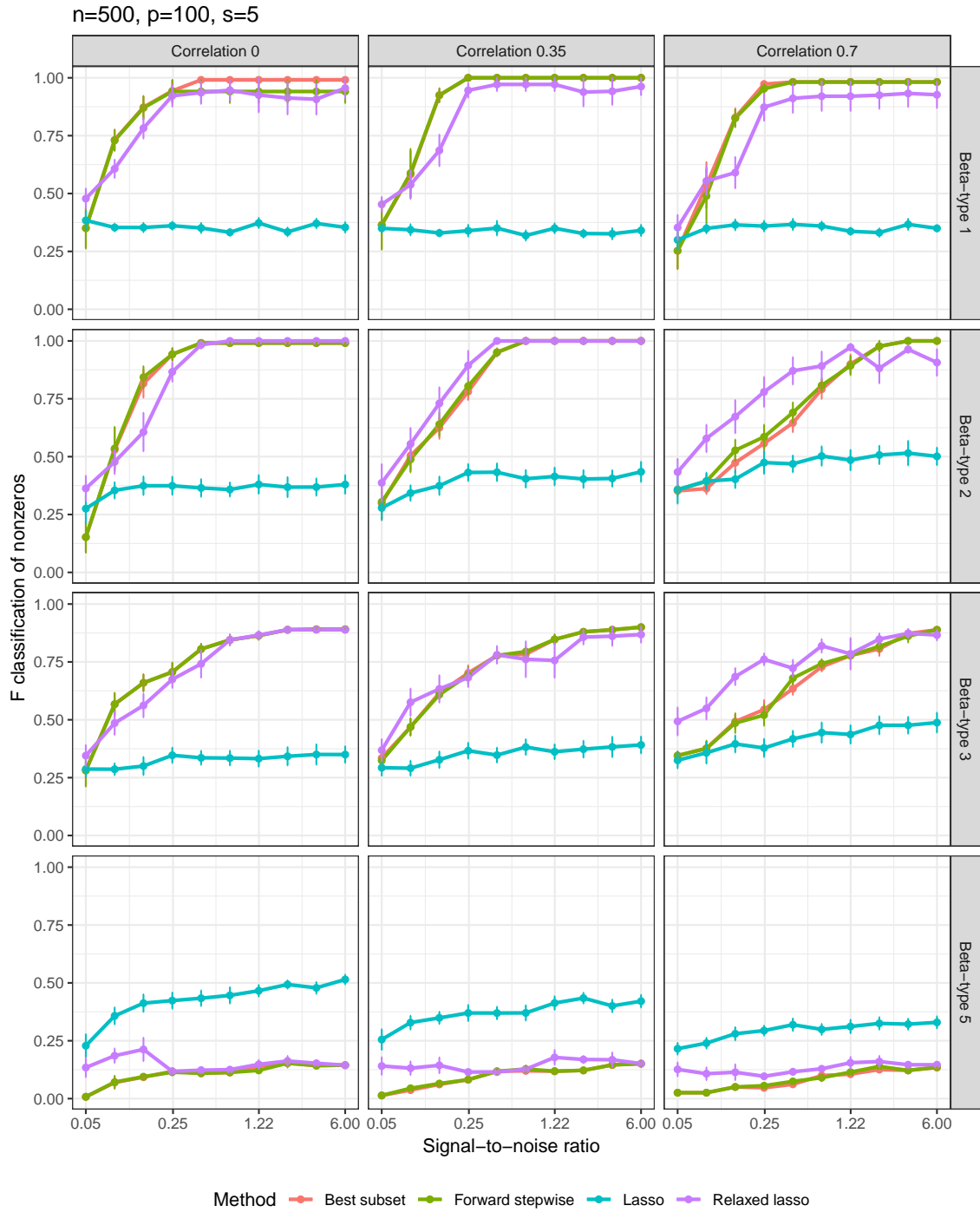
A.2.3 Proportion of variance explained



A.2.4 Number of nonzero coefficients

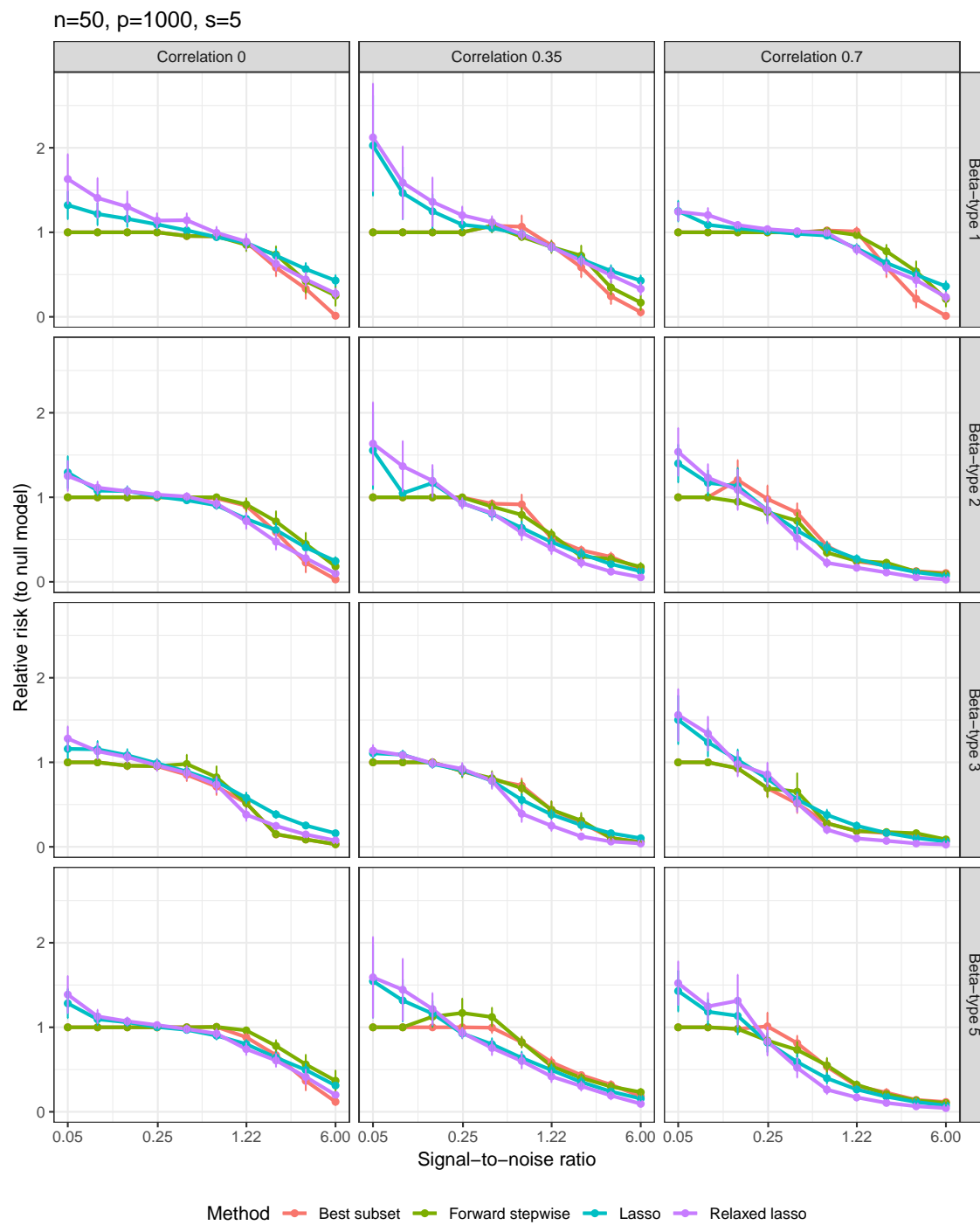


A.2.5 F-score

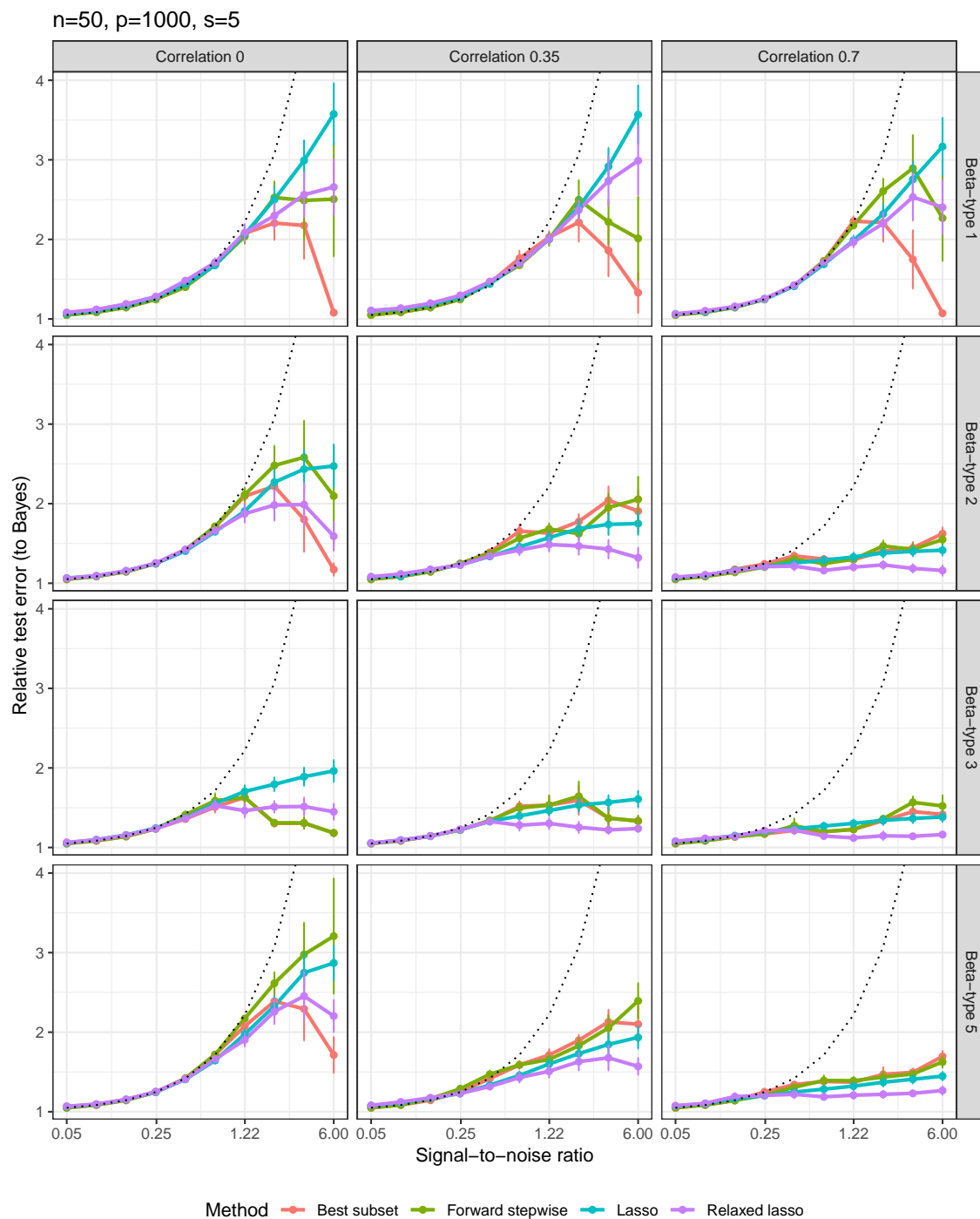


A.3 High-5 setting: $n = 50, p = 1000, s = 5$

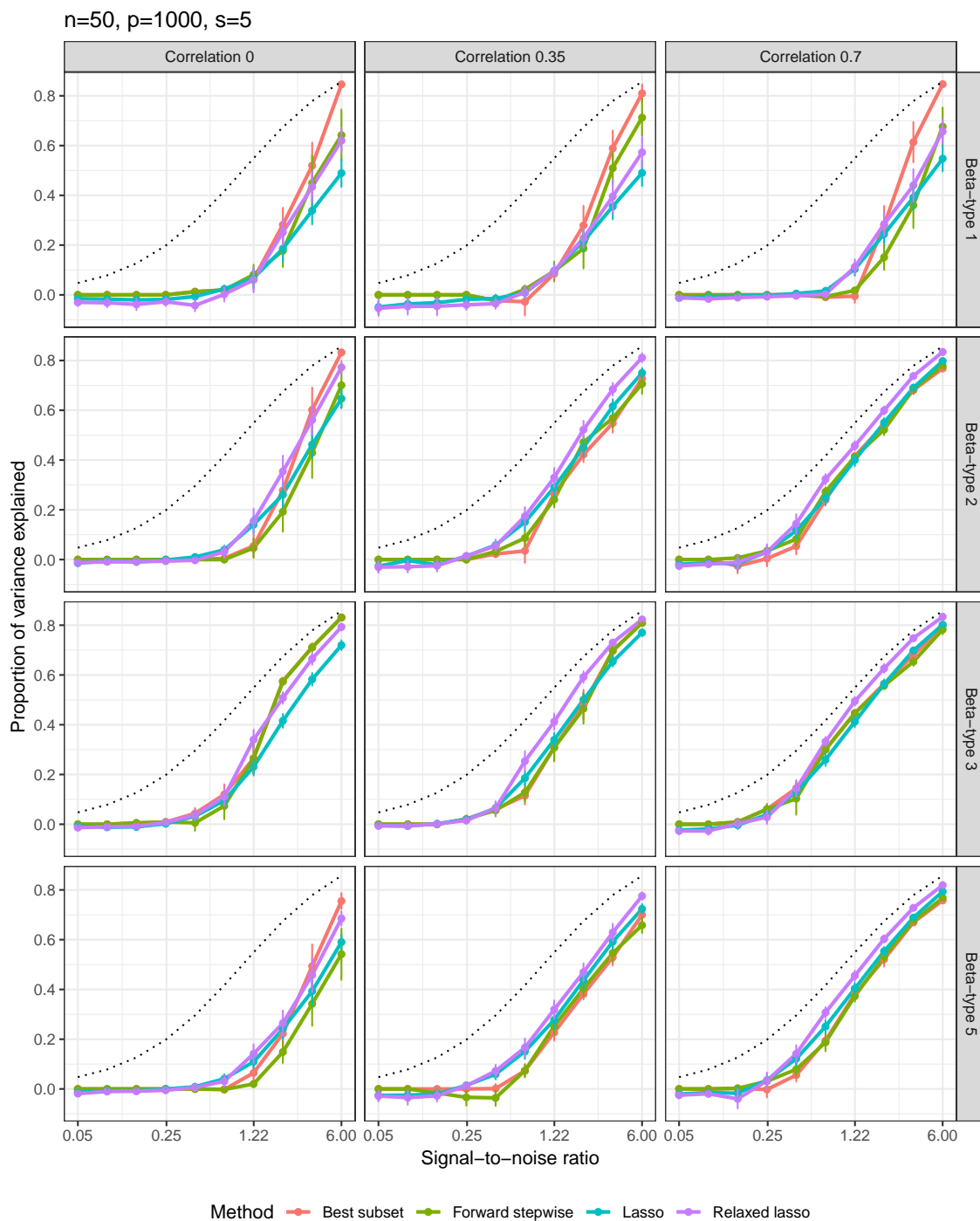
A.3.1 Relative risk (to null model)



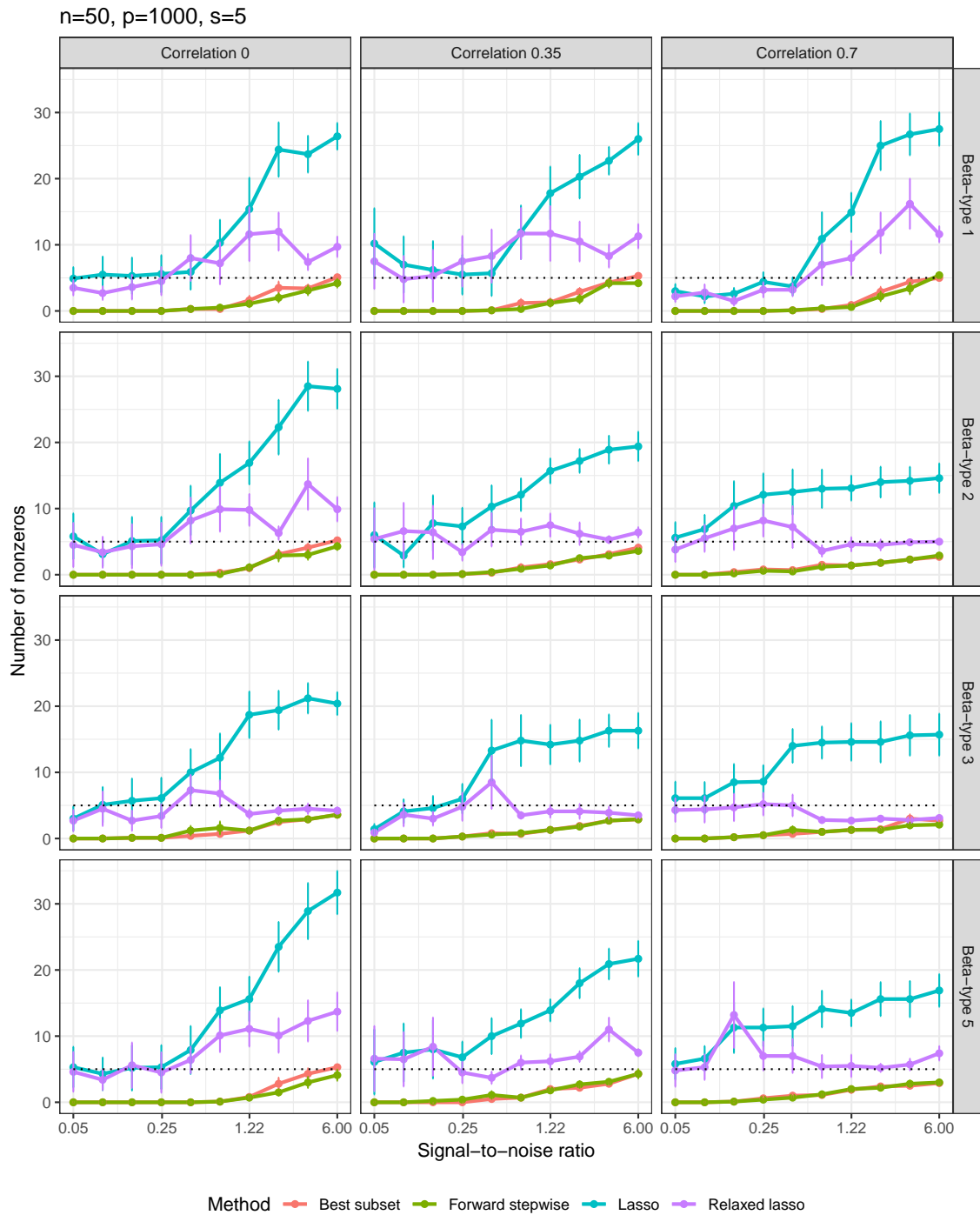
A.3.2 Relative test error (to Bayes)



A.3.3 Proportion of variance explained

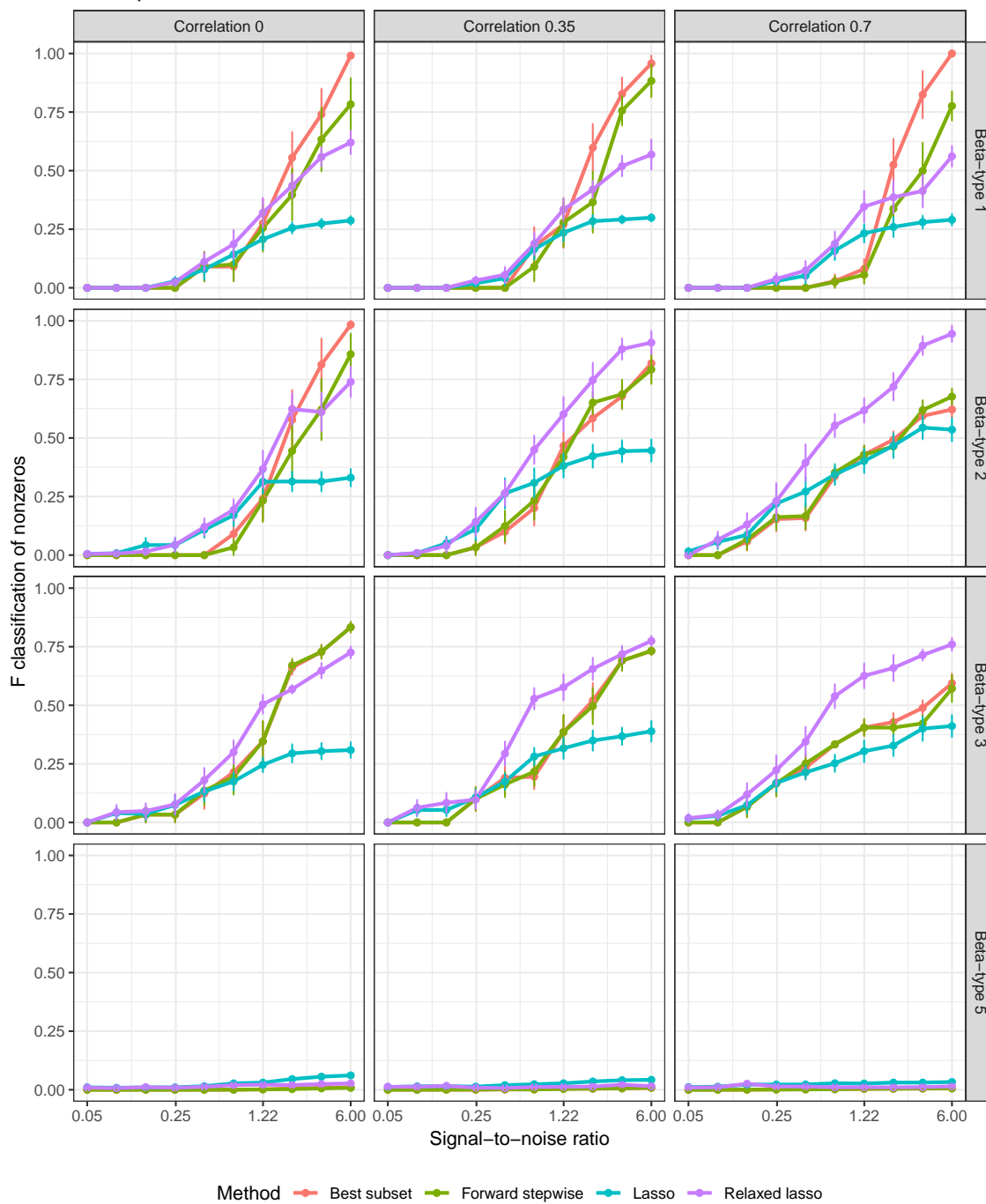


A.3.4 Number of nonzero coefficients



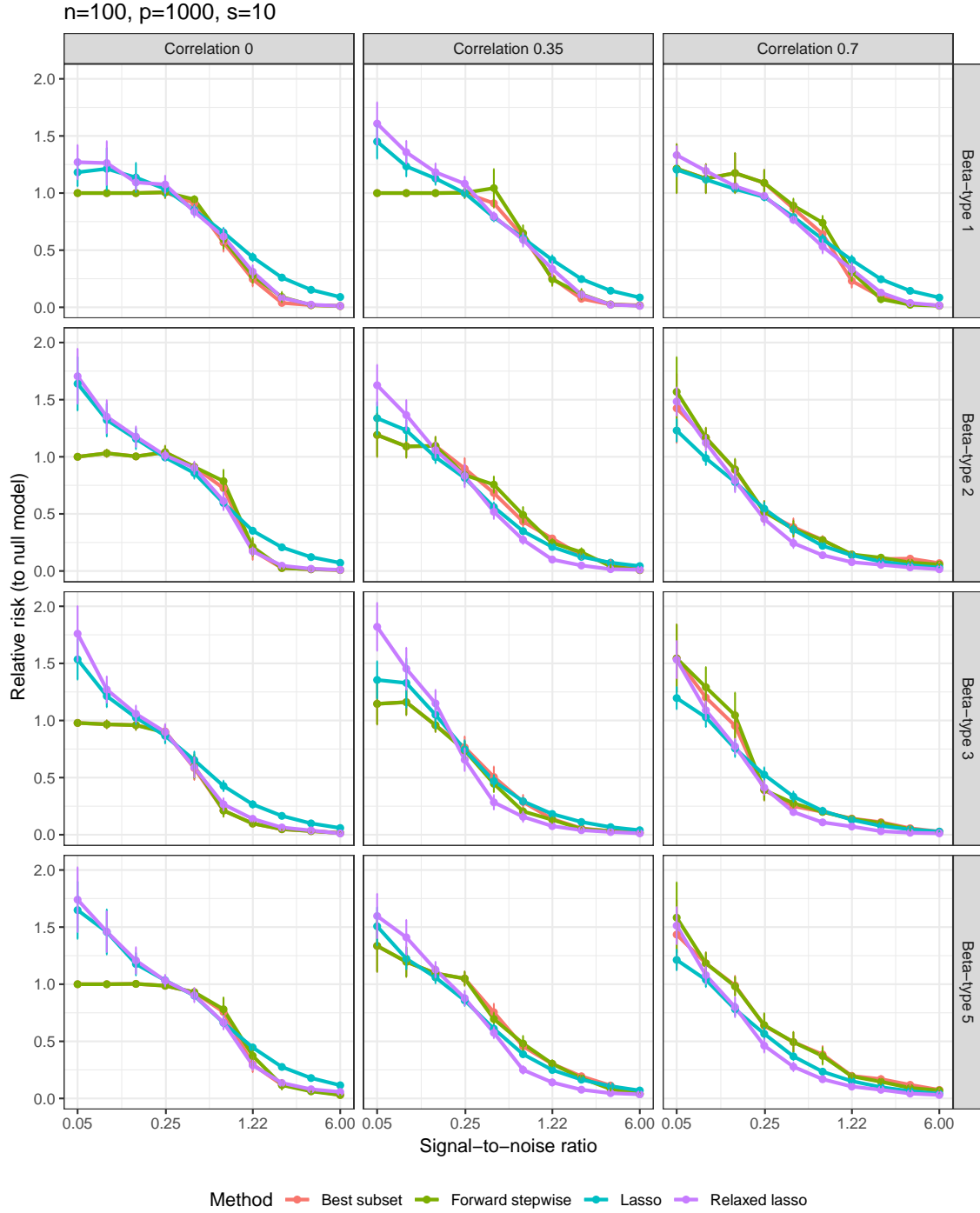
A.3.5 F-score

$n=50, p=1000, s=5$

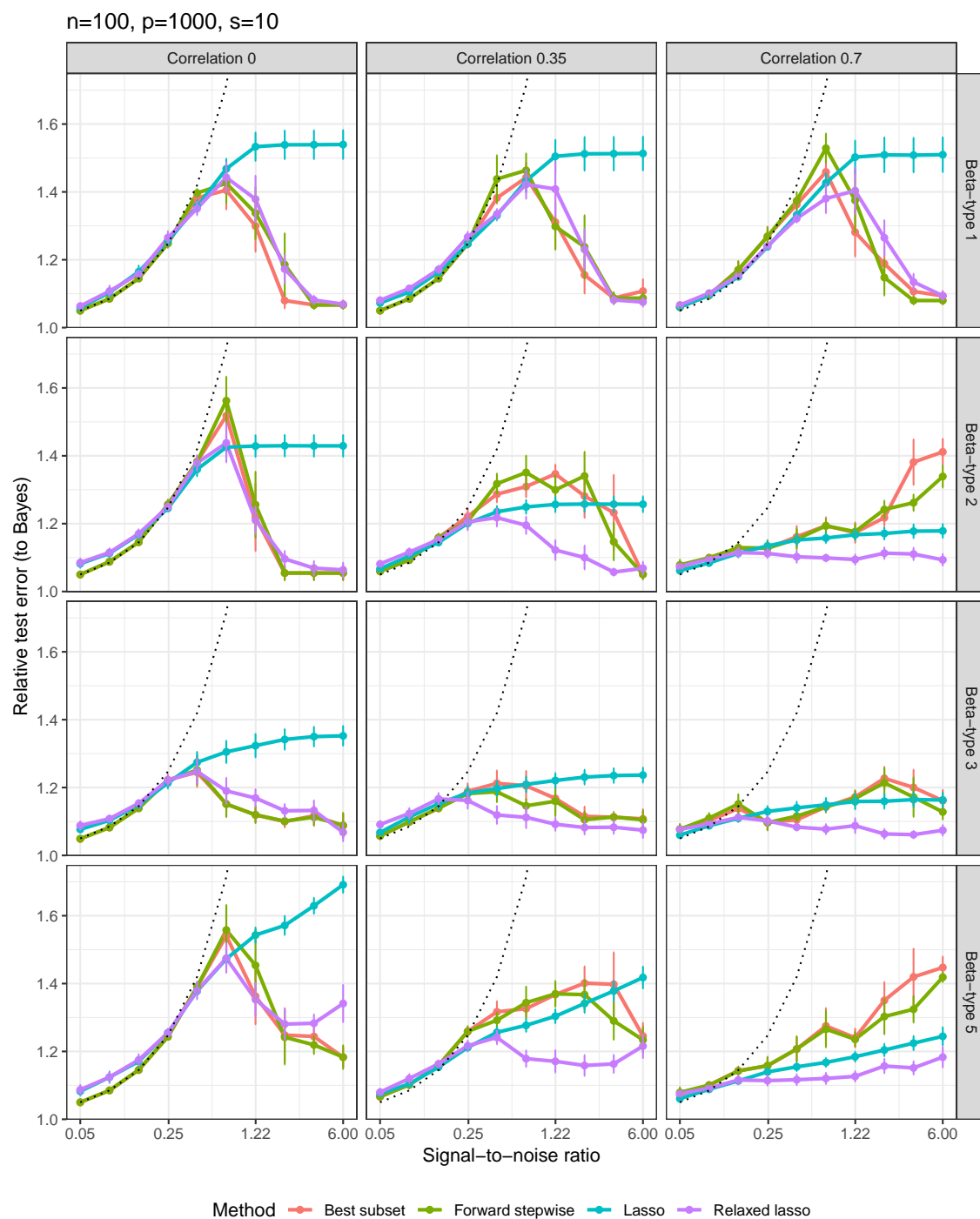


A.4 High-10 setting: $n = 100, p = 1000, s = 10$

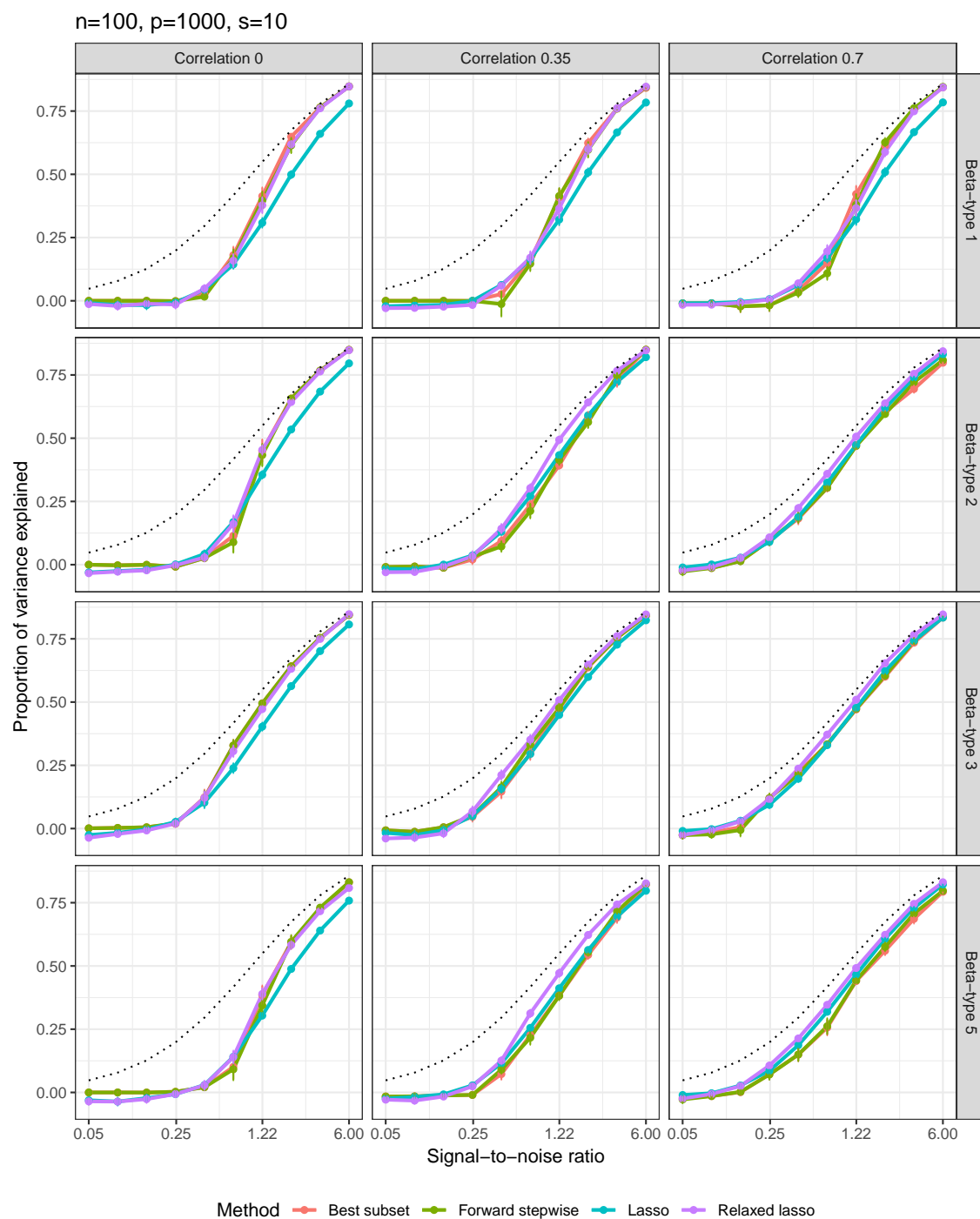
A.4.1 Relative risk (to null model)



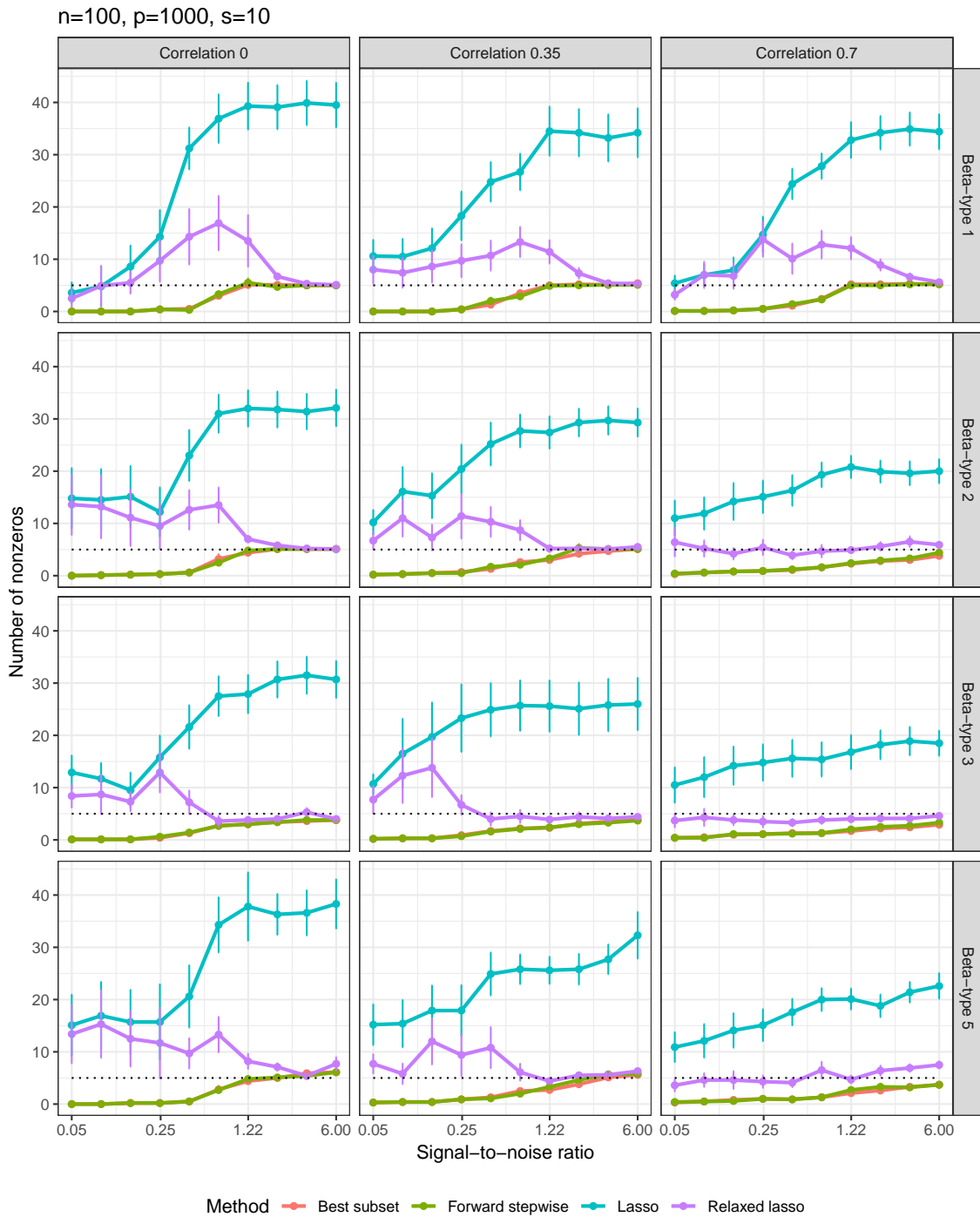
A.4.2 Relative test error (to Bayes)



A.4.3 Proportion of variance explained

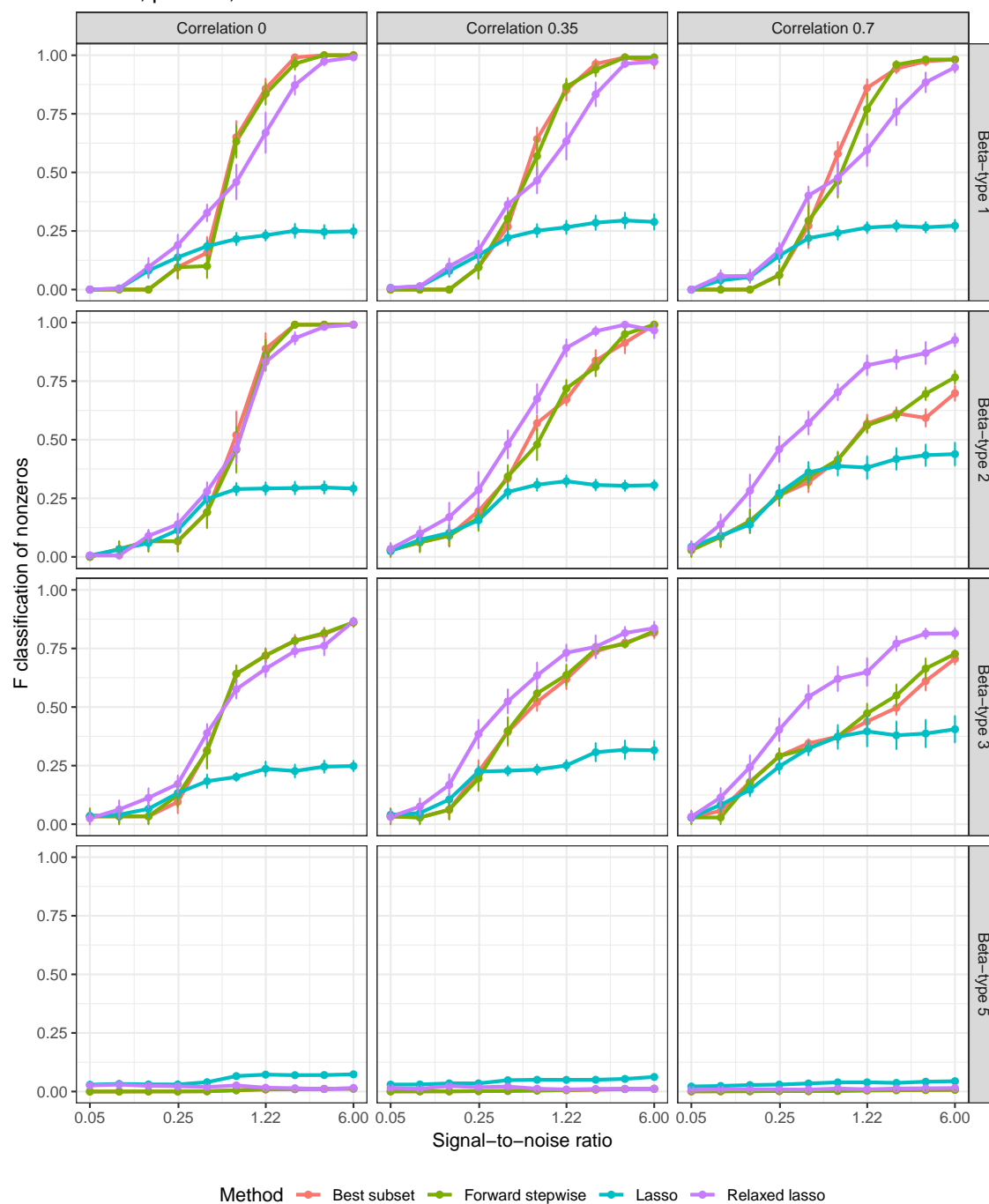


A.4.4 Number of nonzero coefficients



A.4.5 F-score

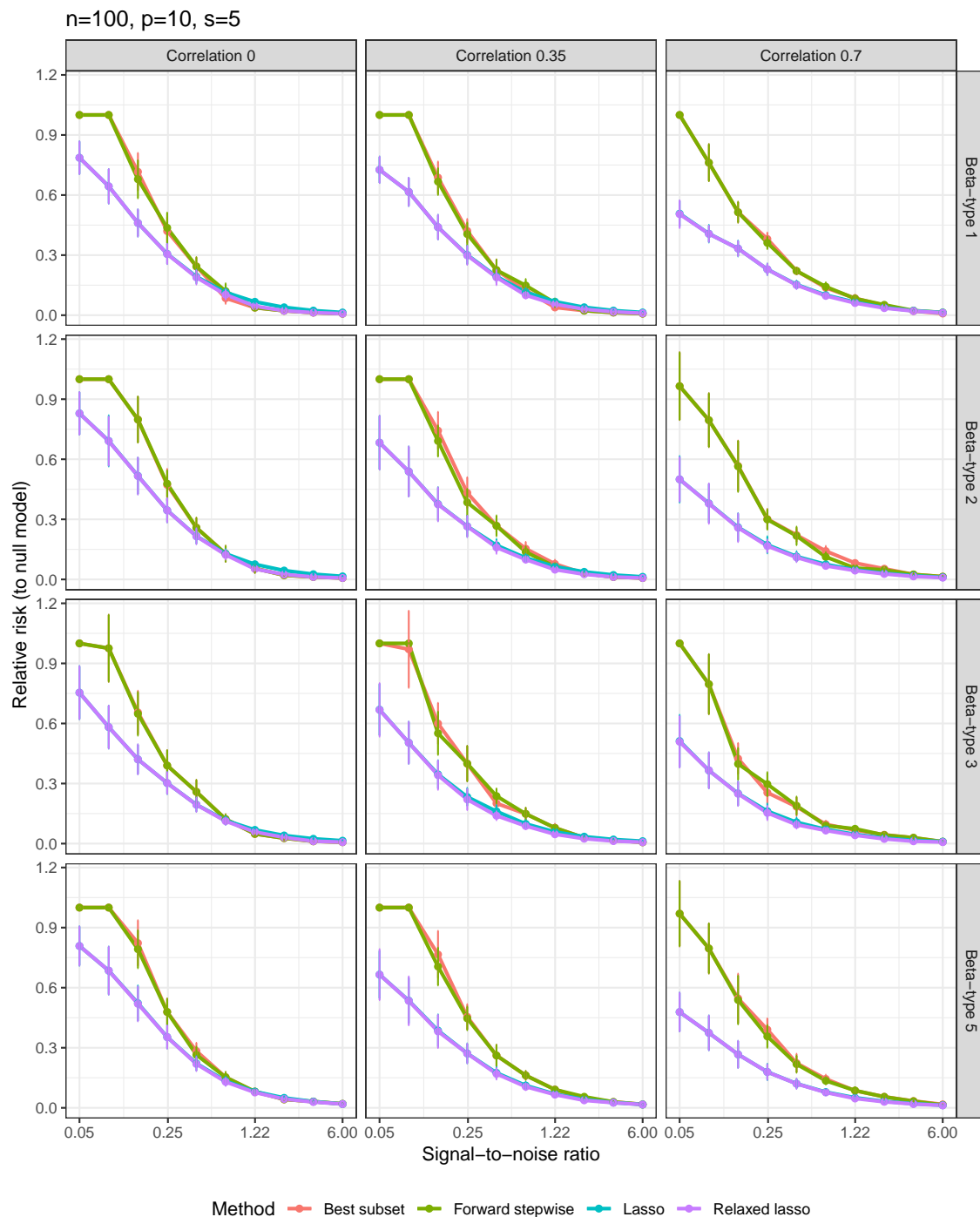
$n=100, p=1000, s=10$



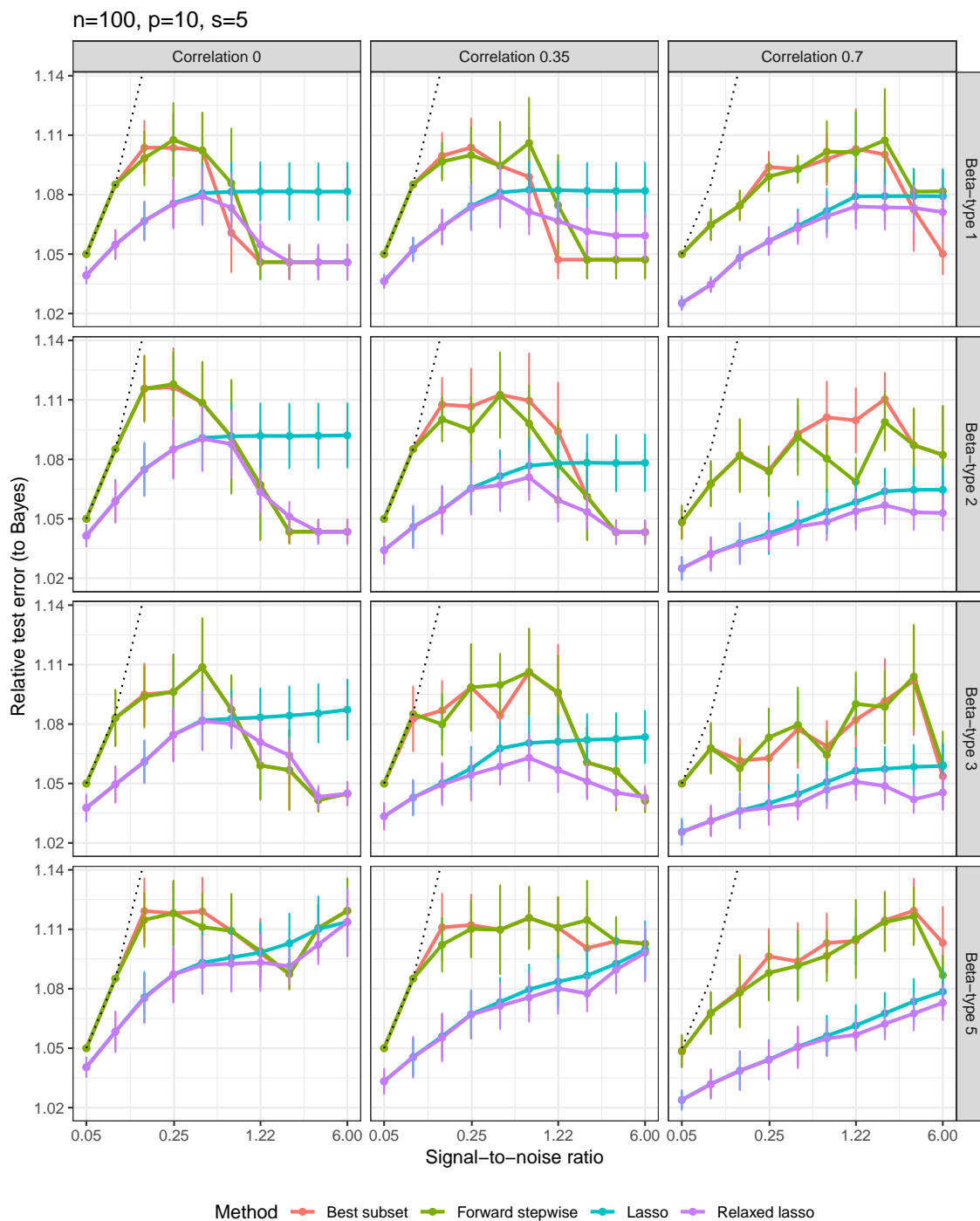
B Oracle tuning

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

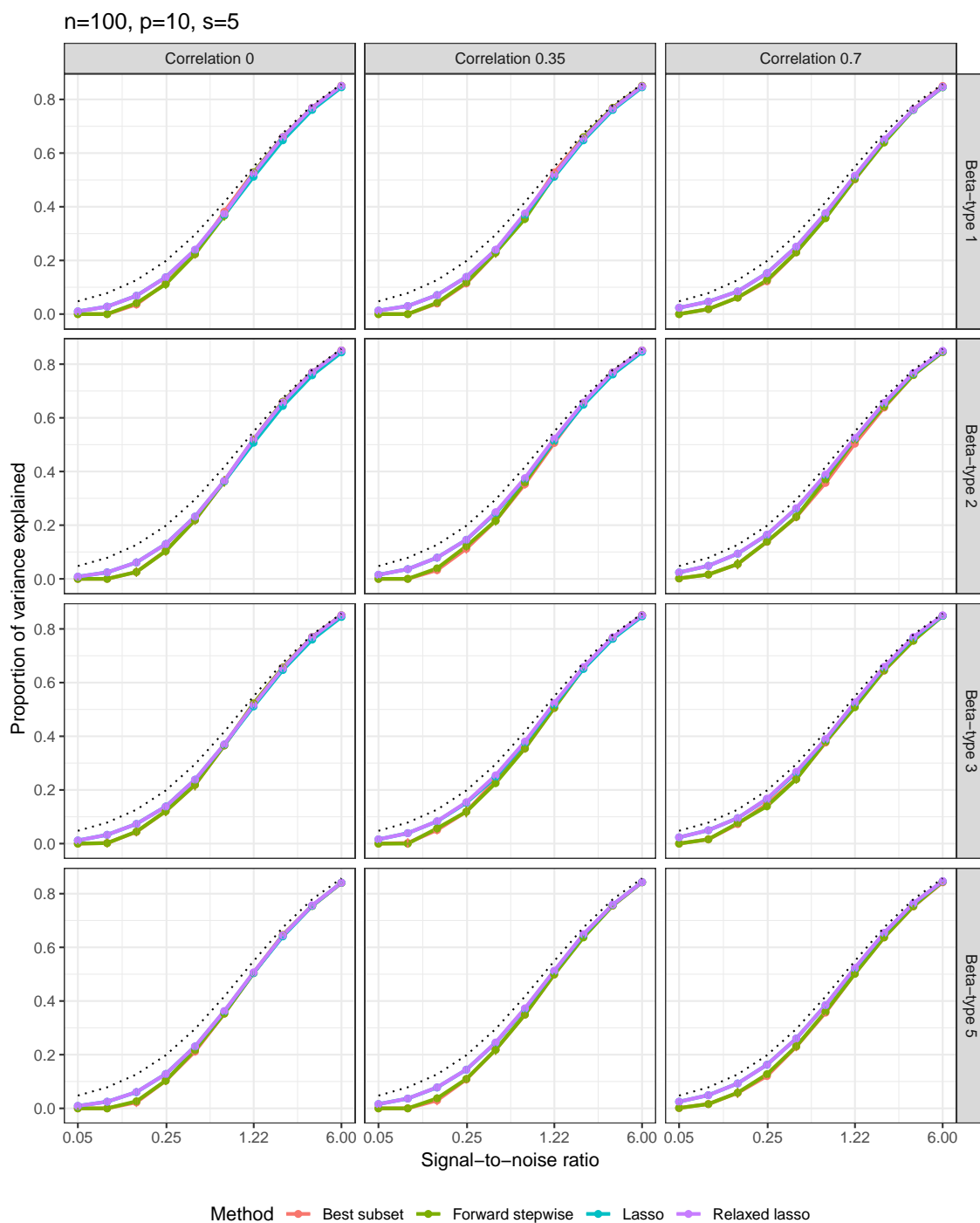
B.1.1 Relative risk (to null model)



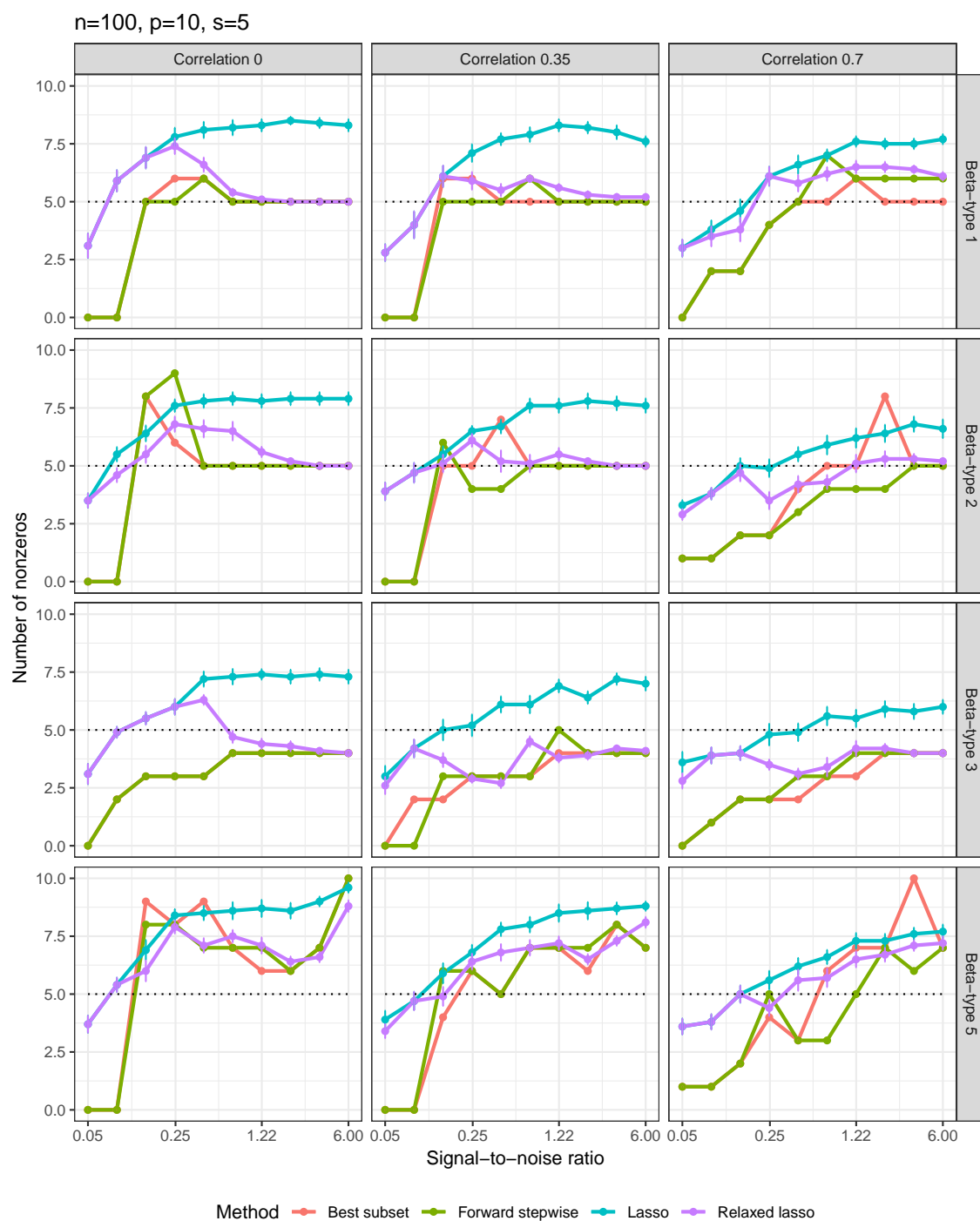
B.1.2 Relative test error (to Bayes)



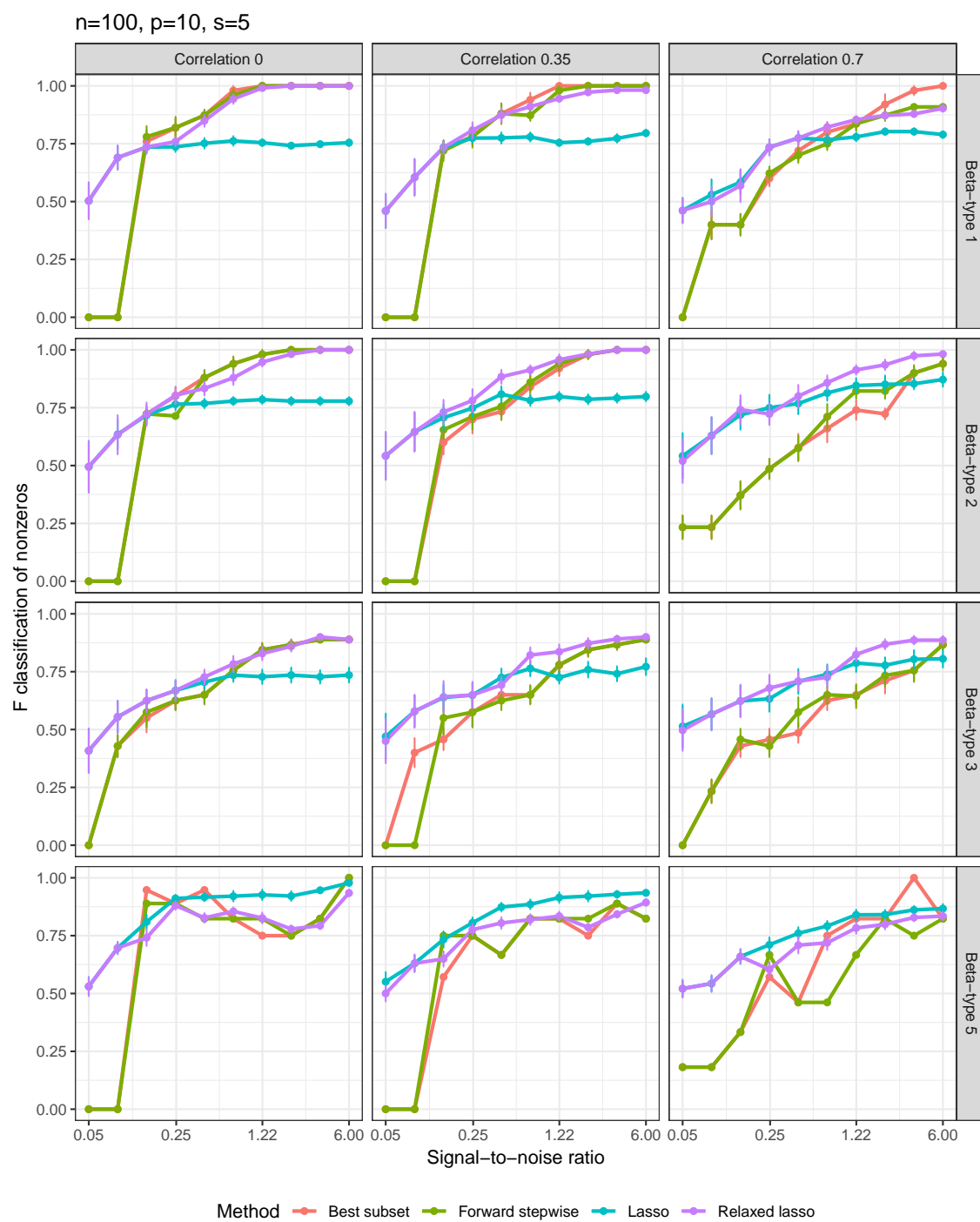
B.1.3 Proportion of variance explained



B.1.4 Number of nonzero coefficients

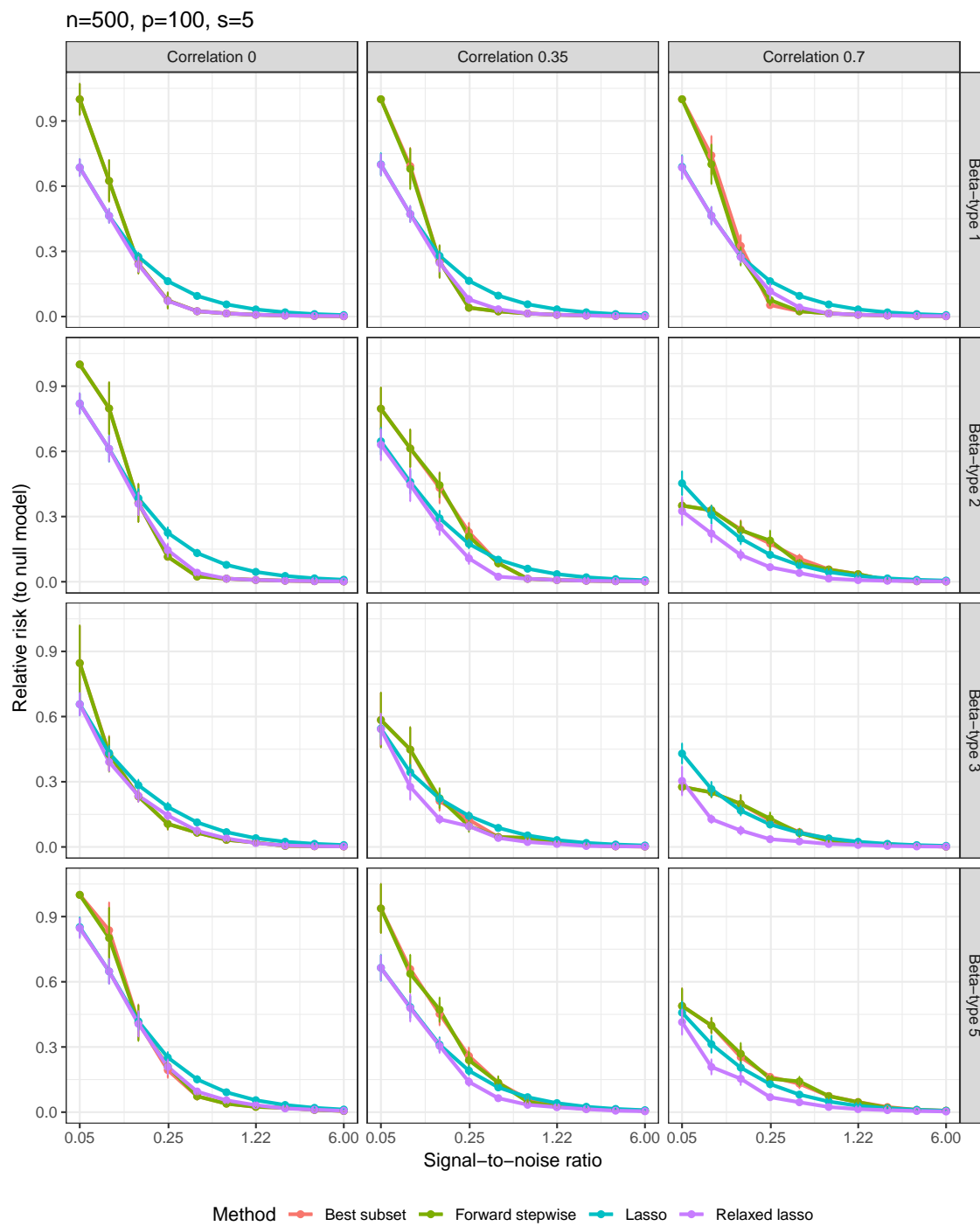


B.1.5 F-score

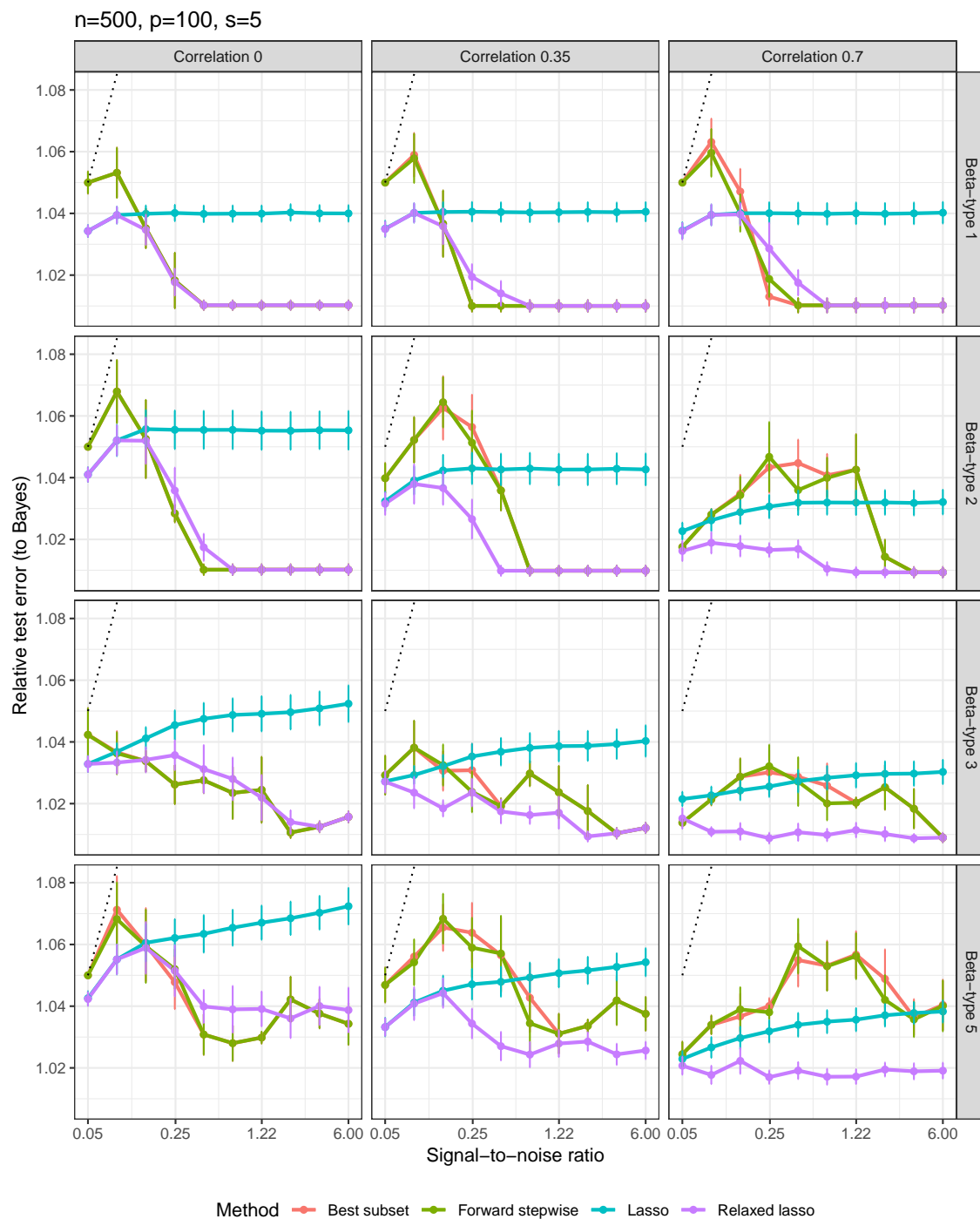


B.2 Medium setting: $n = 500, p = 100, s = 5$

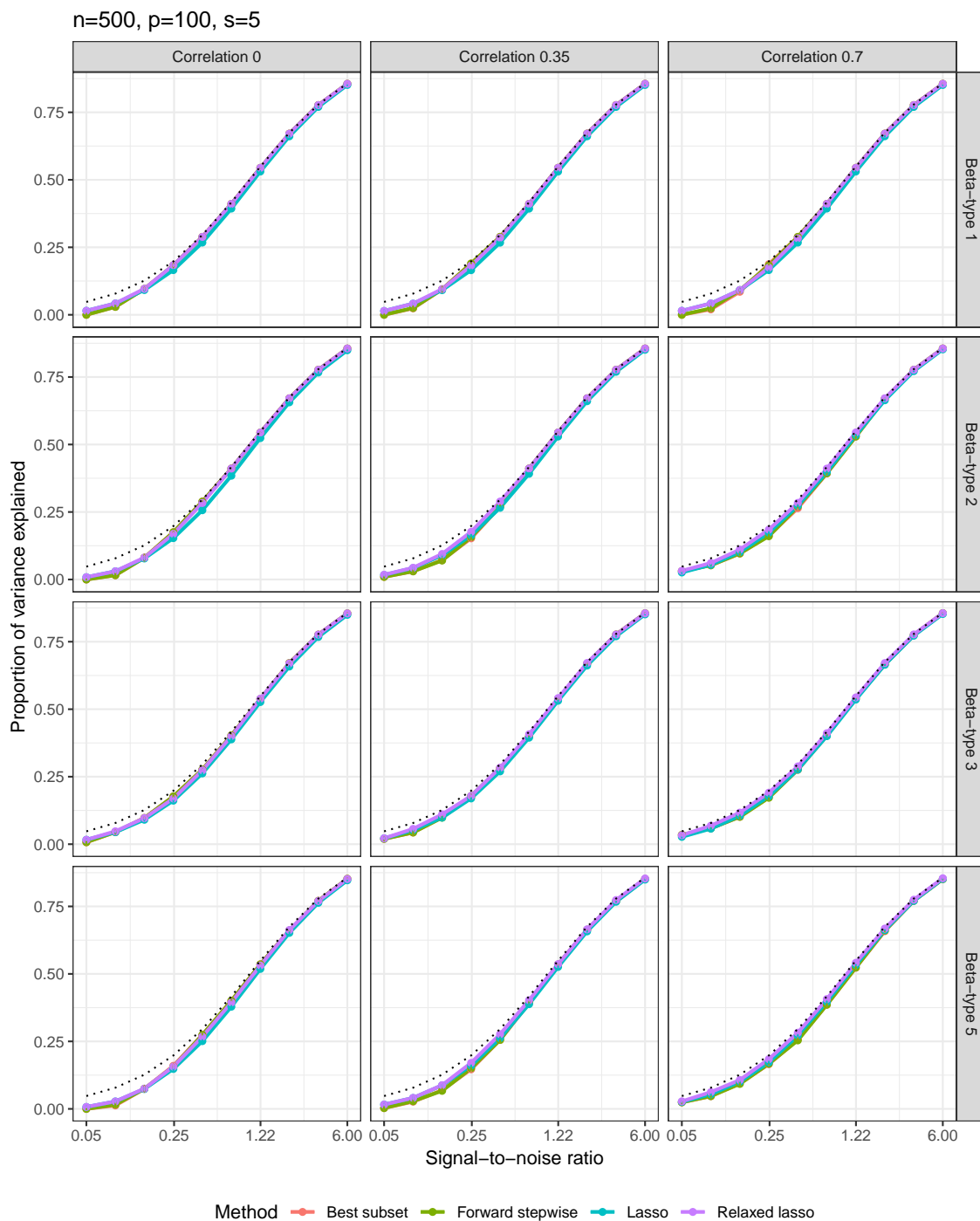
B.2.1 Relative risk (to null model)



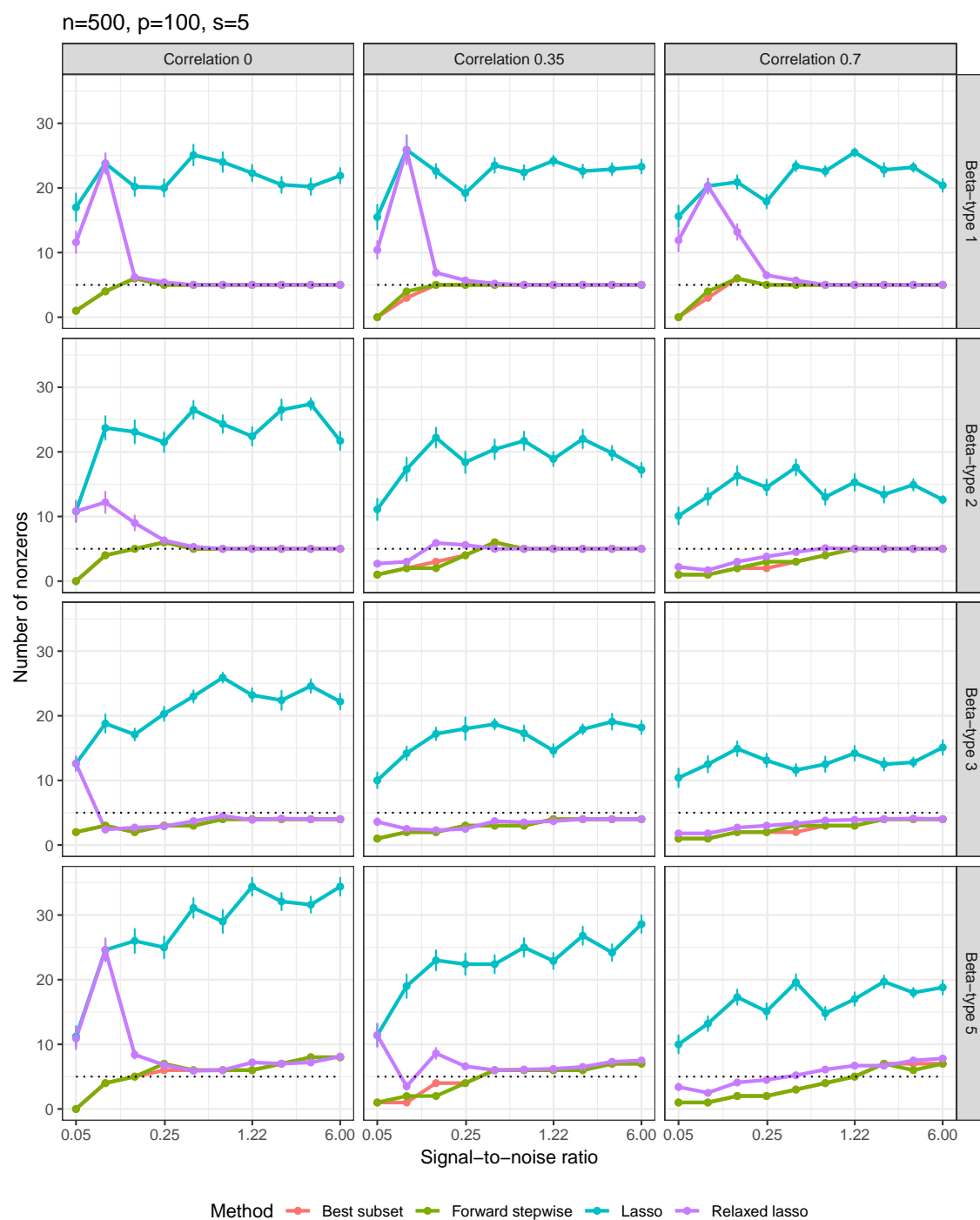
B.2.2 Relative test error (to Bayes)



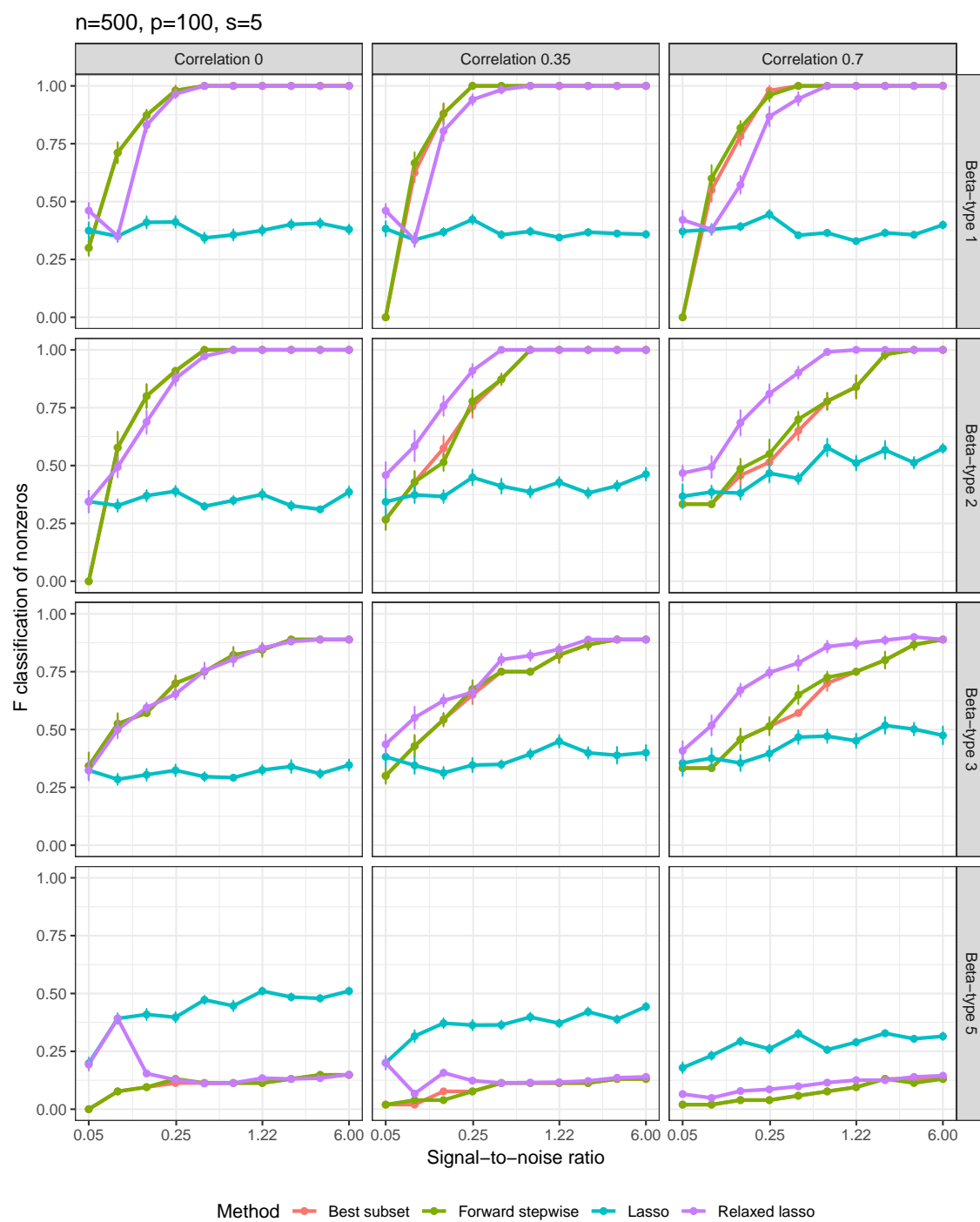
B.2.3 Proportion of variance explained



B.2.4 Number of nonzero coefficients

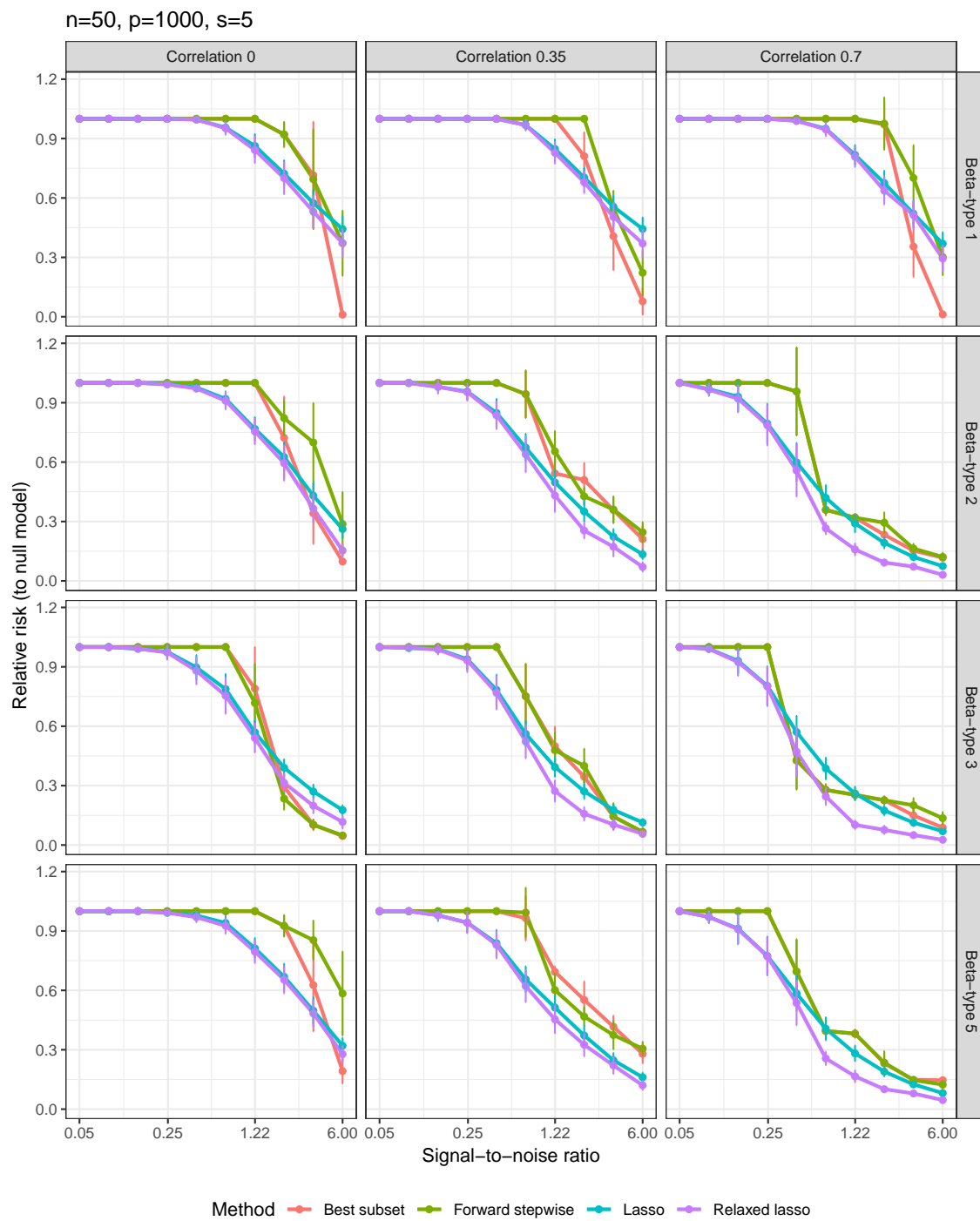


B.2.5 F-score

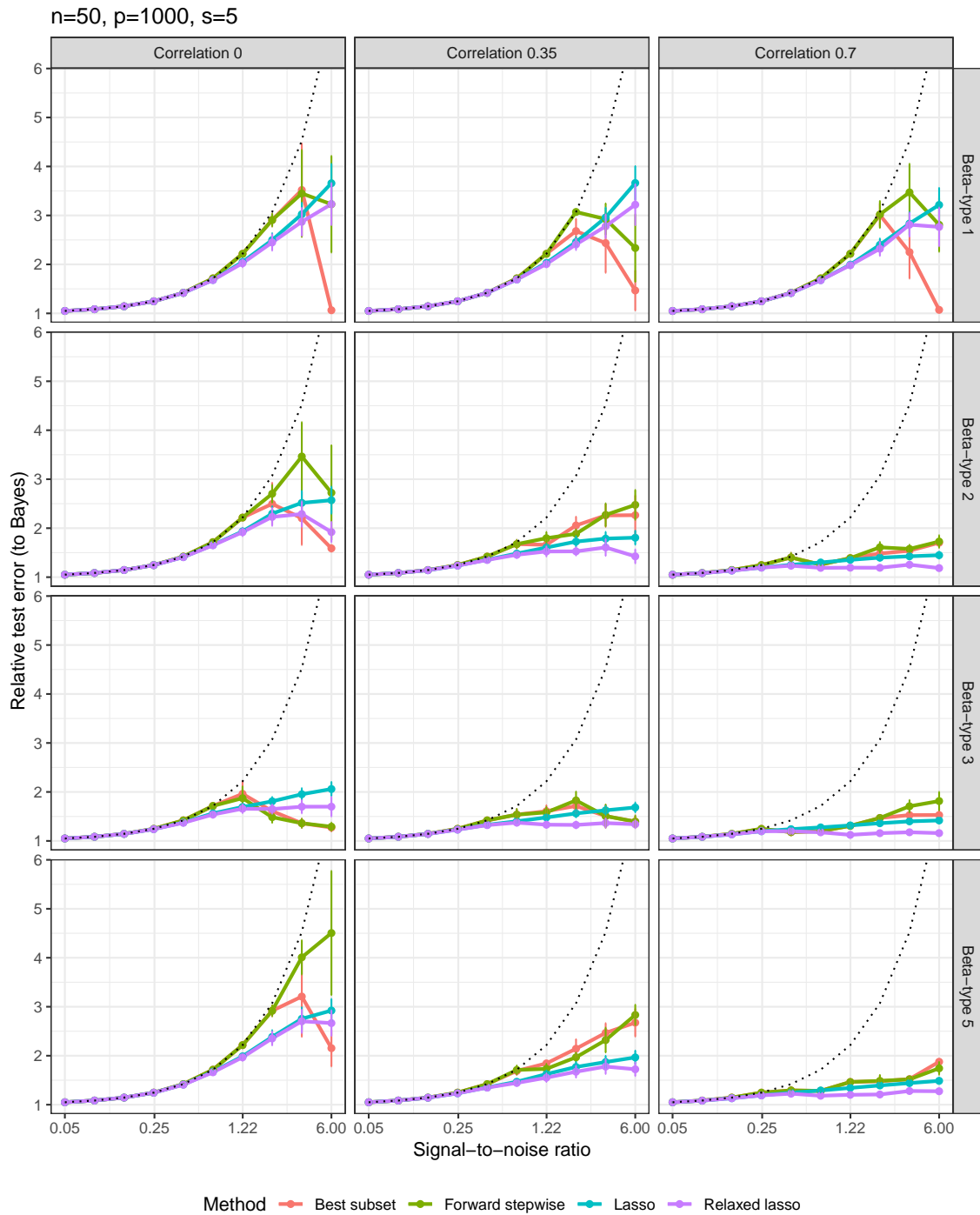


B.3 High-5 setting: $n = 50, p = 1000, s = 5$

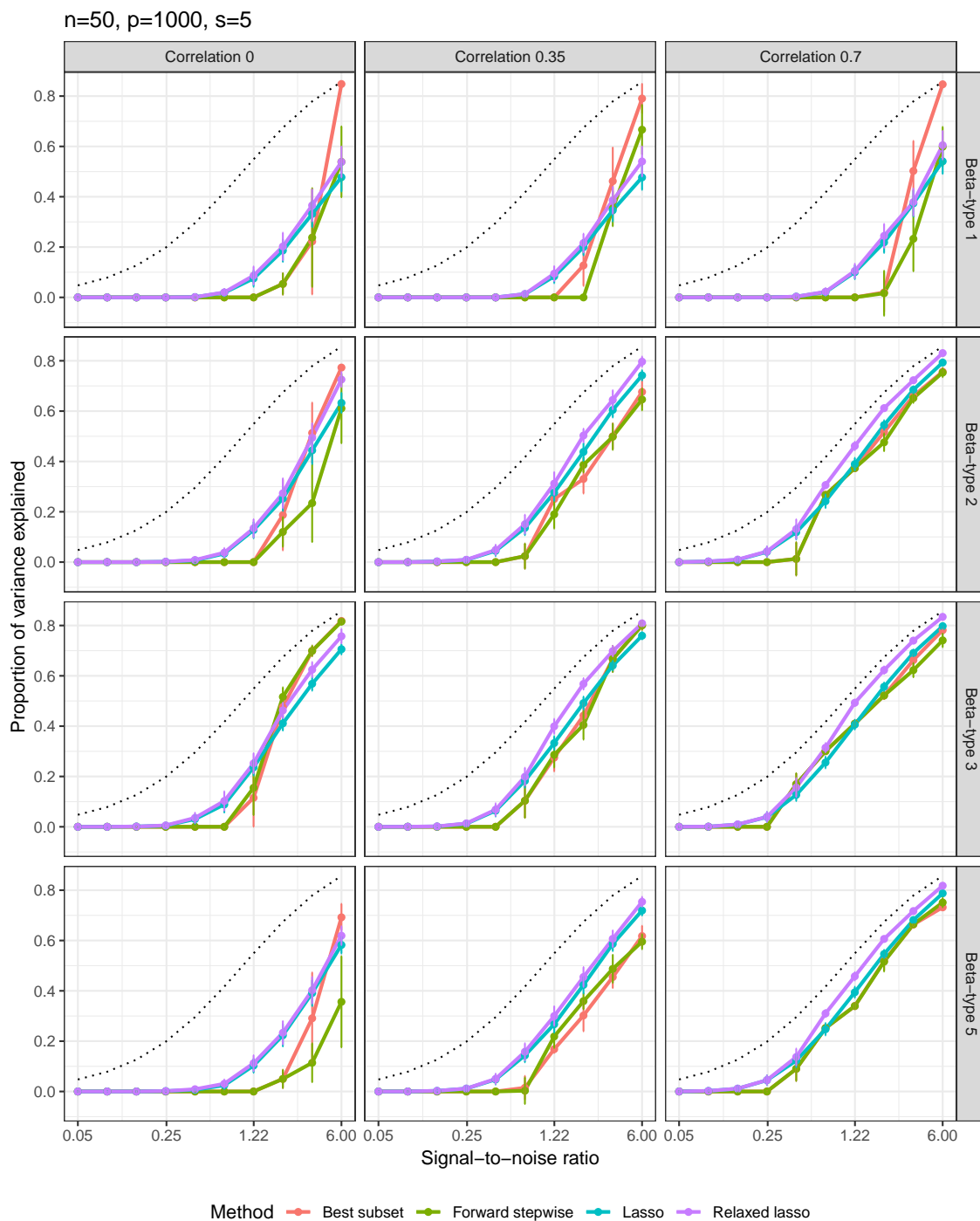
B.3.1 Relative risk (to null model)



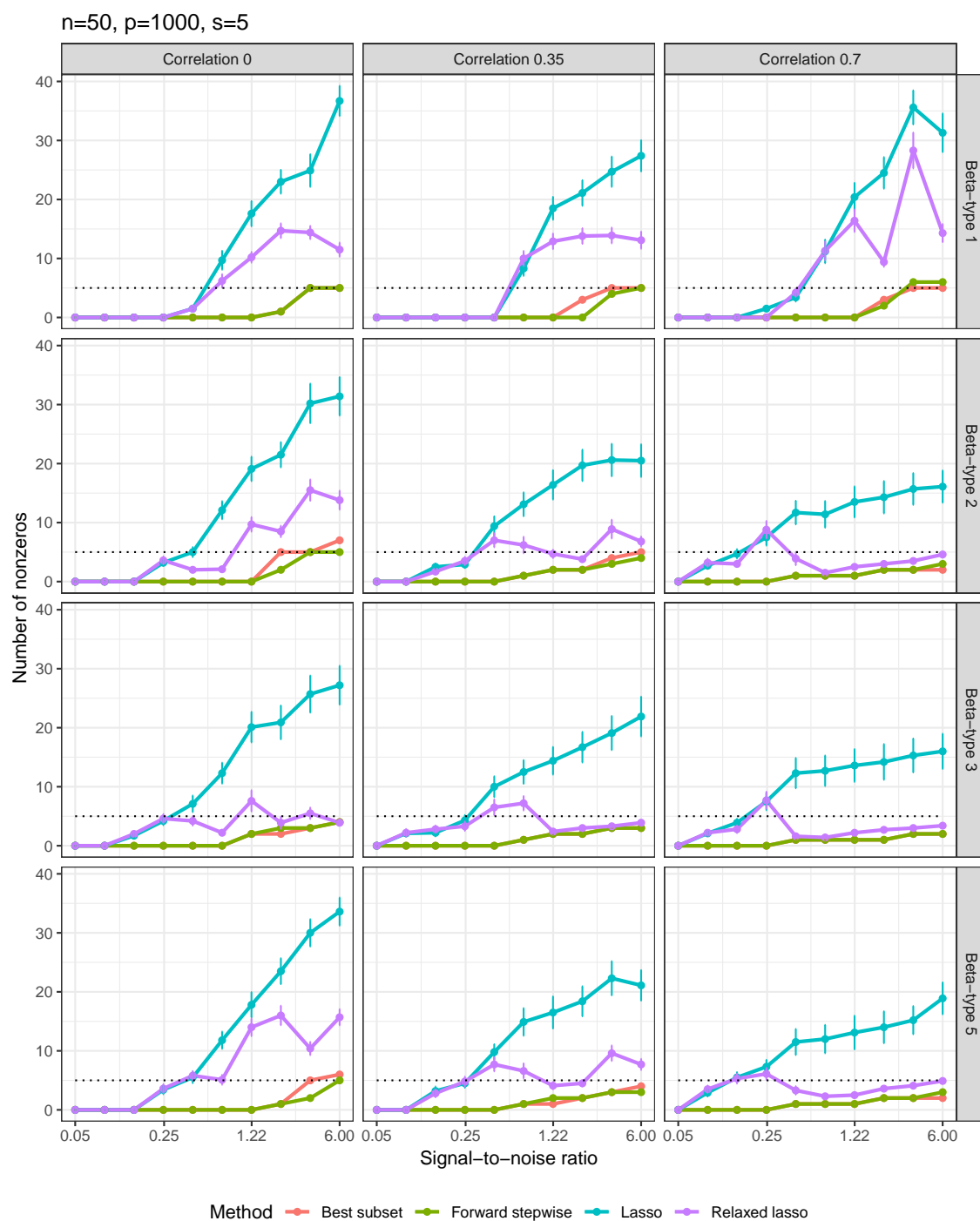
B.3.2 Relative test error (to Bayes)



B.3.3 Proportion of variance explained

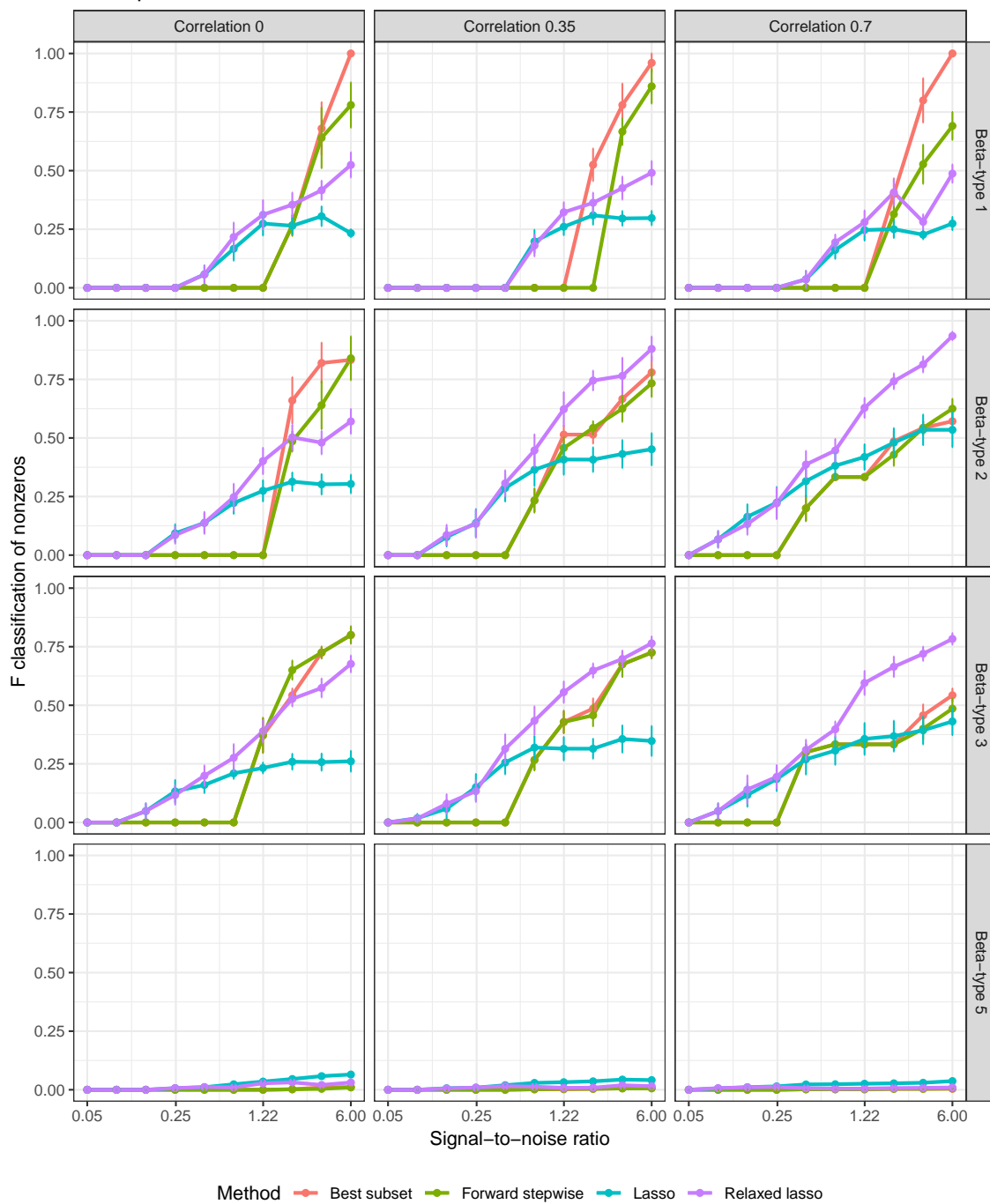


B.3.4 Number of nonzero coefficients



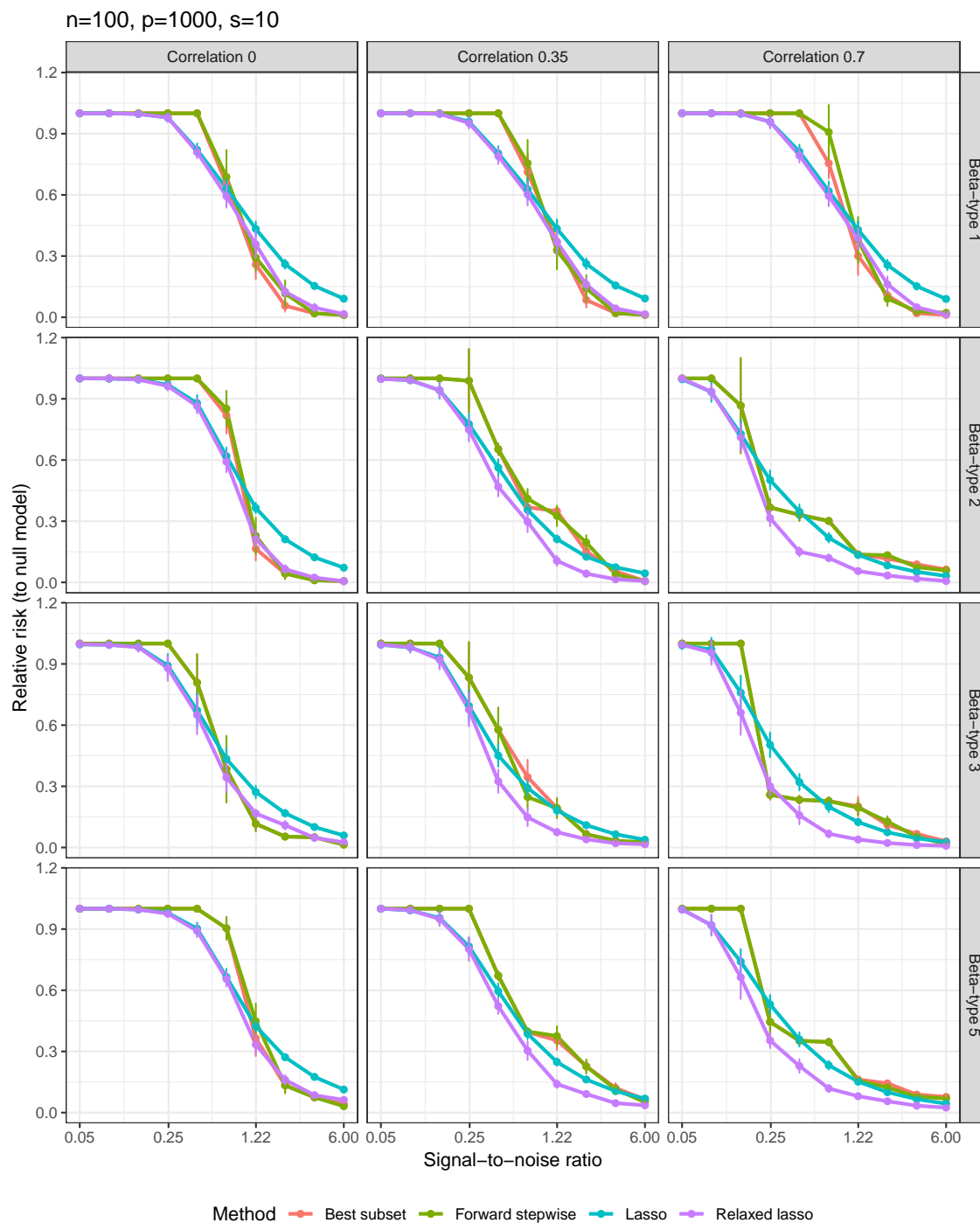
B.3.5 F-score

$n=50, p=1000, s=5$

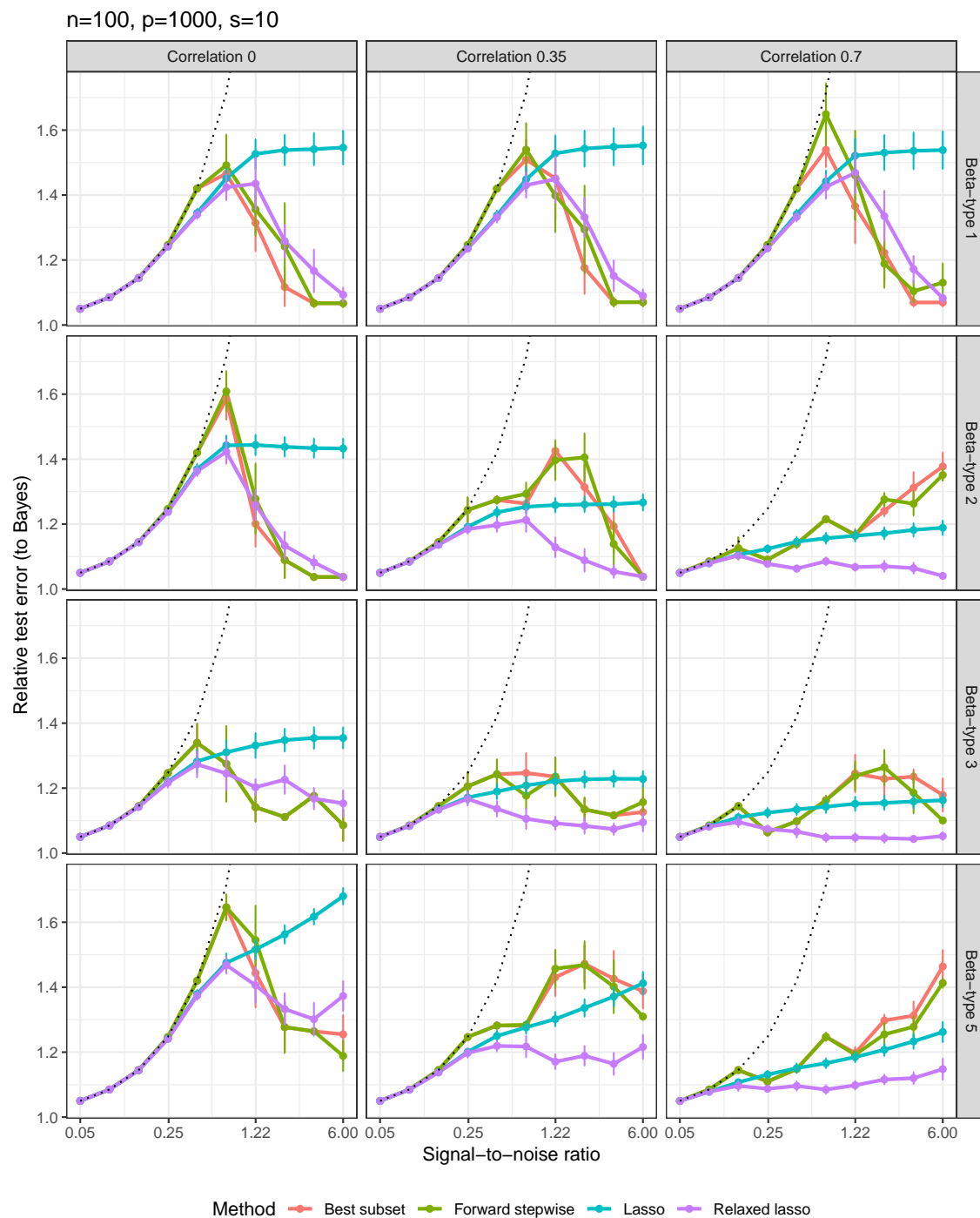


B.4 High-10 setting: $n = 100, p = 1000, s = 10$

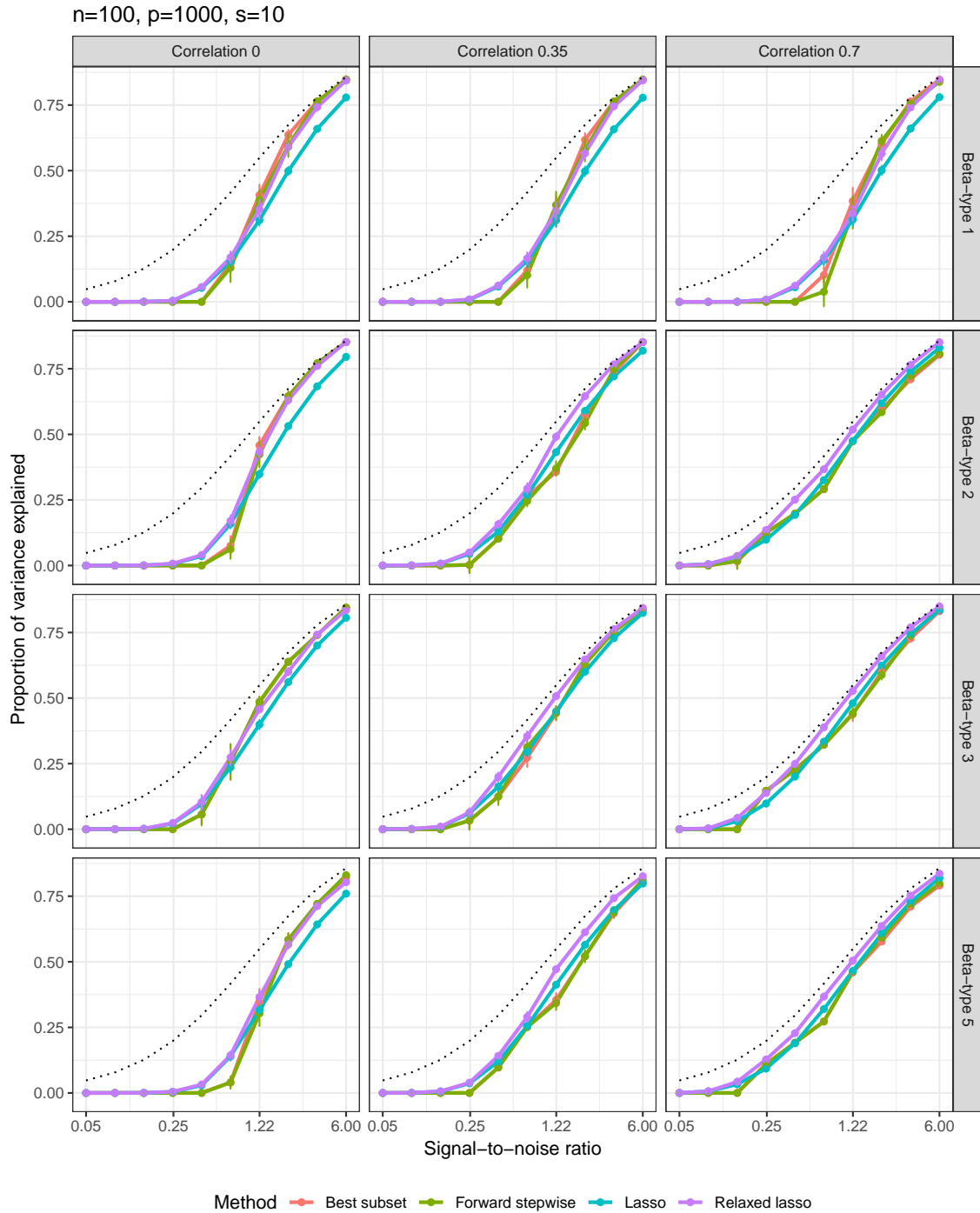
B.4.1 Relative risk (to null model)



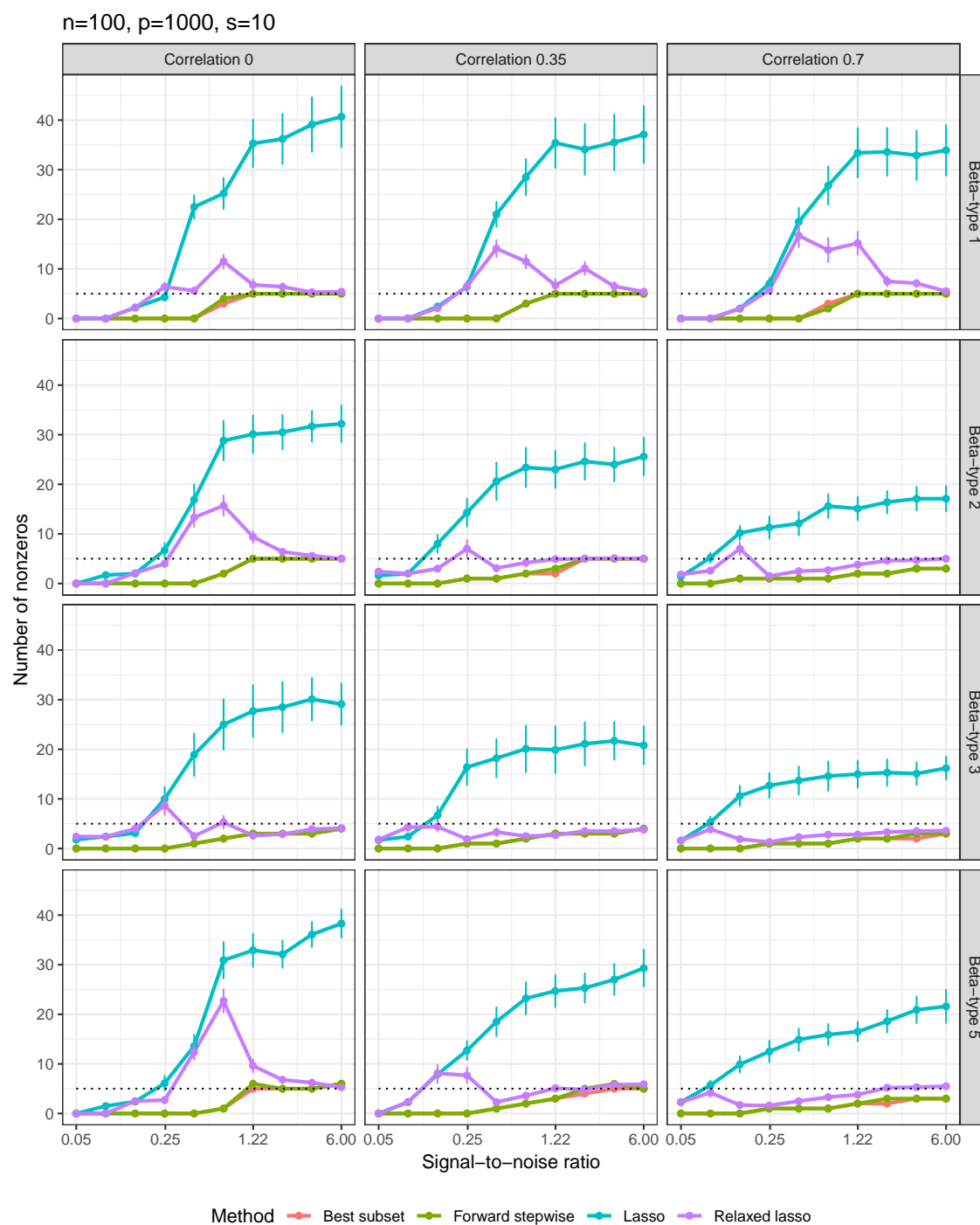
B.4.2 Relative test error (to Bayes)



B.4.3 Proportion of variance explained

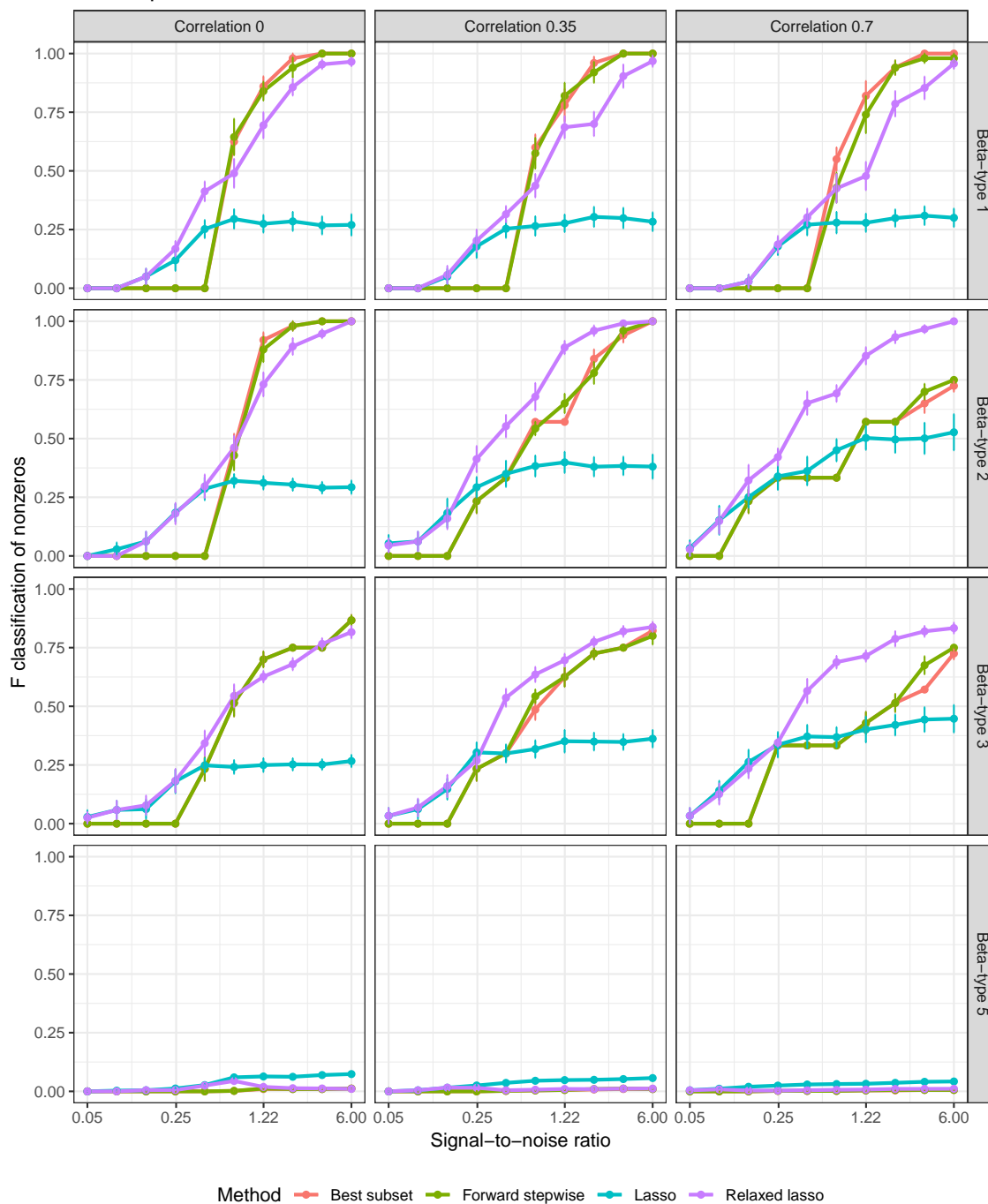


B.4.4 Number of nonzero coefficients



B.4.5 F-score

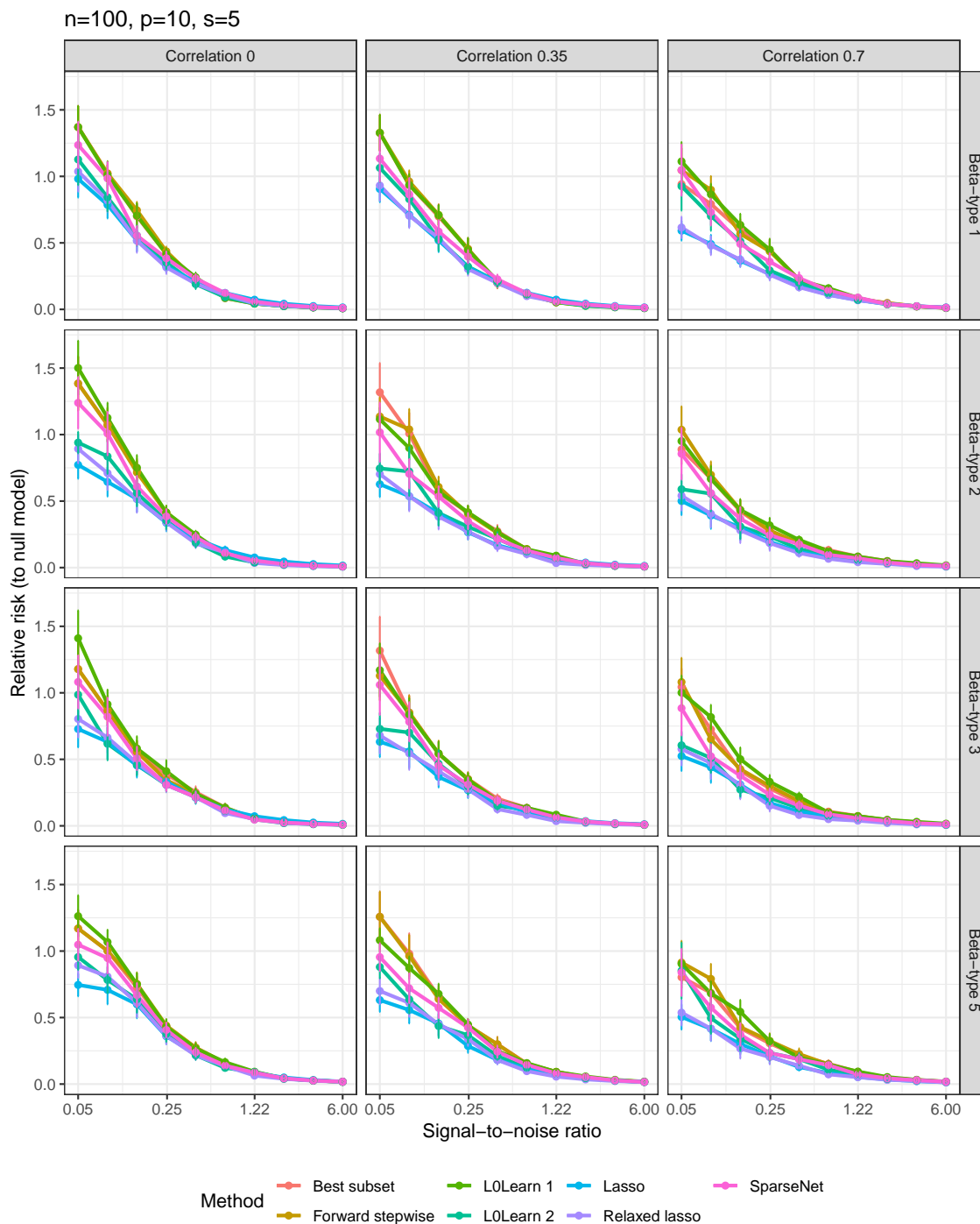
$n=100, p=1000, s=10$



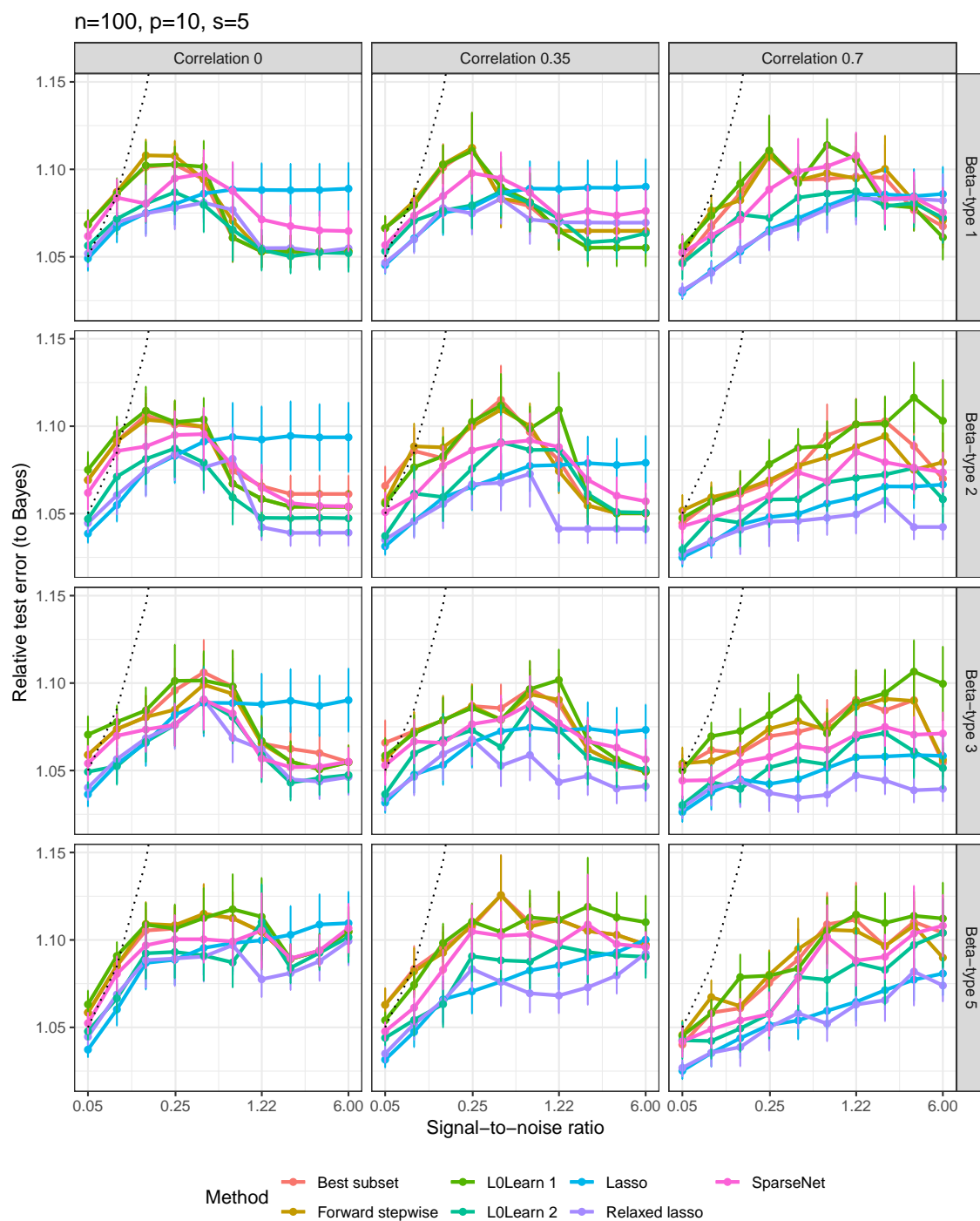
C Validation tuning: more methods

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

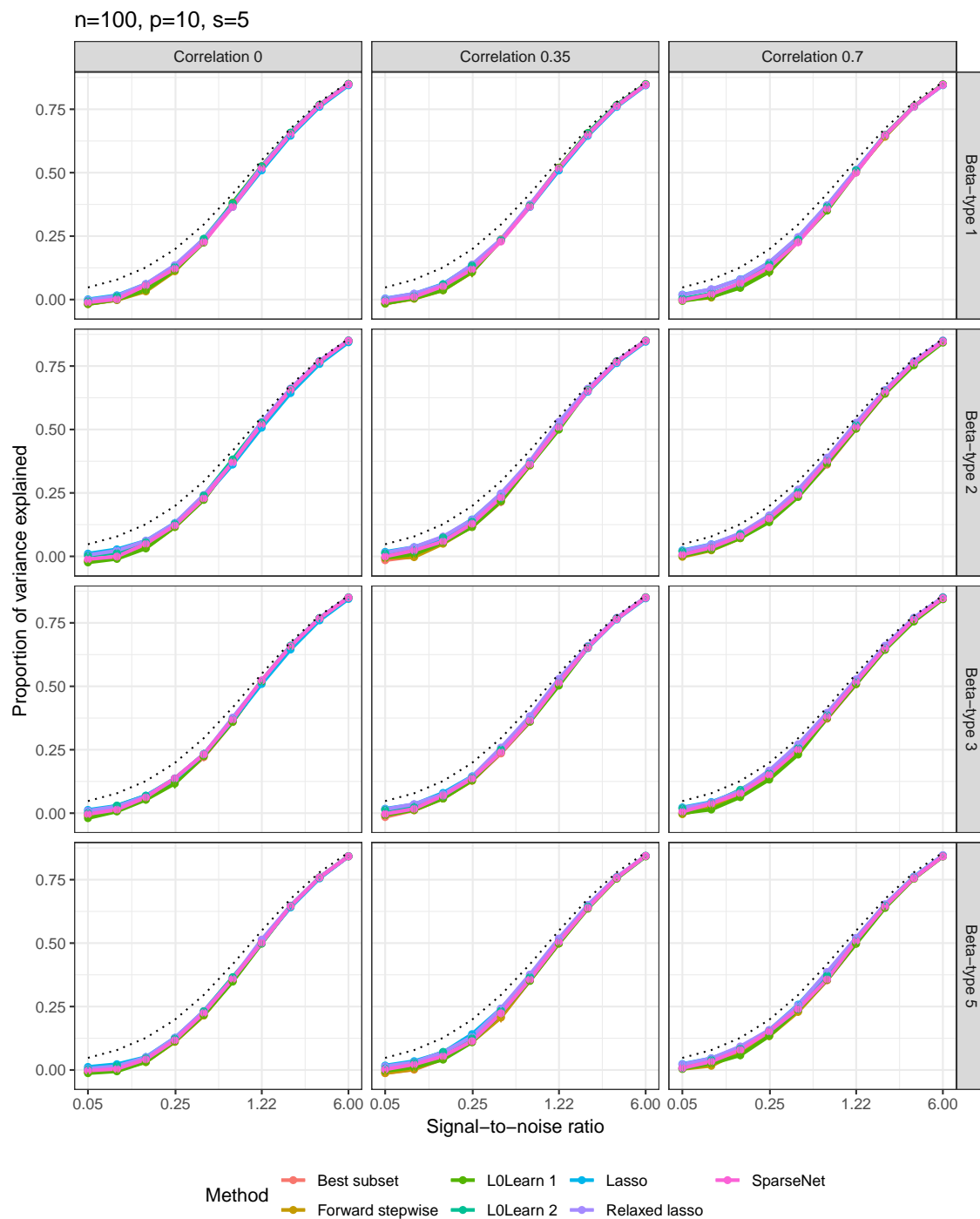
C.1.1 Relative risk (to null model)



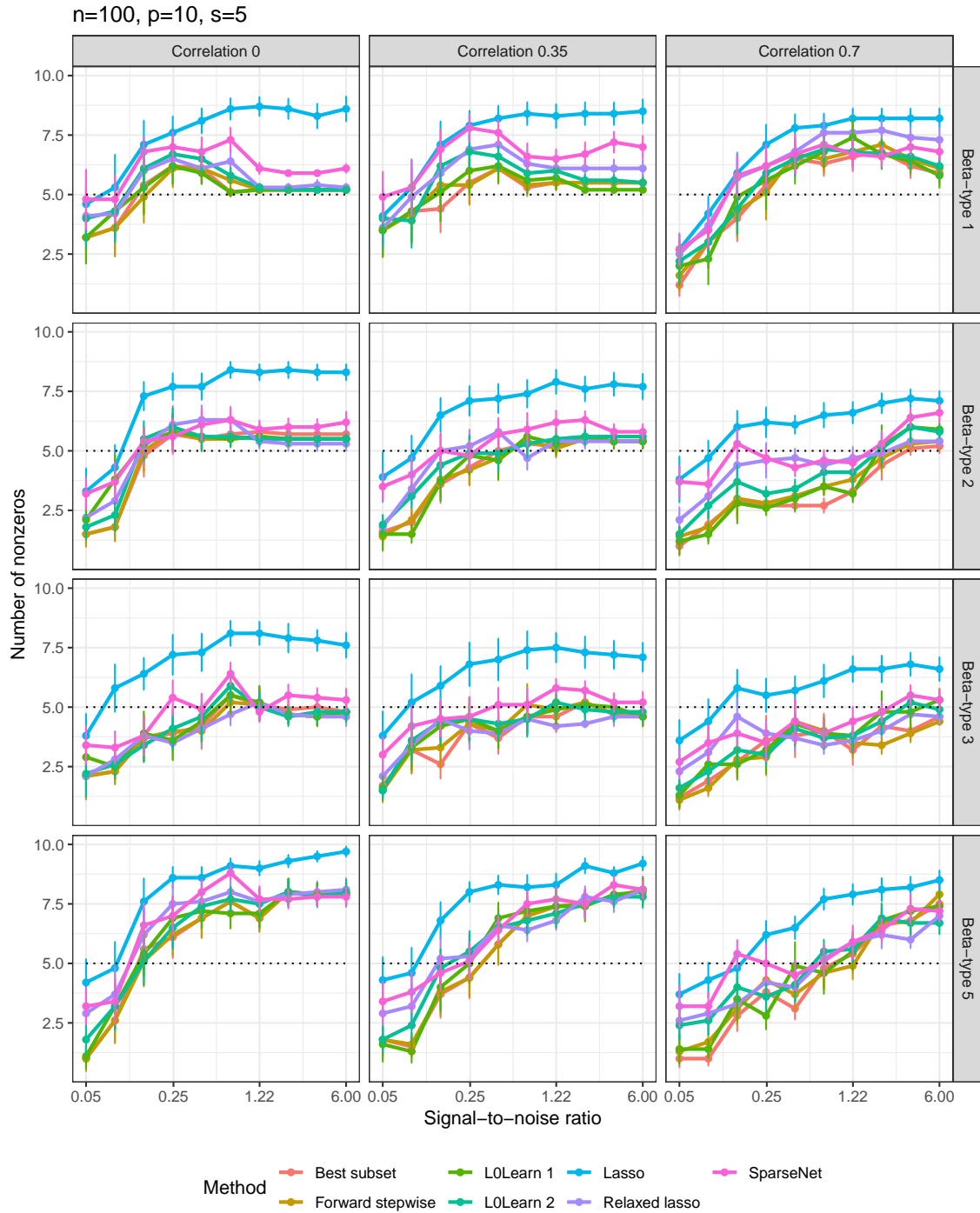
C.1.2 Relative test error (to Bayes)



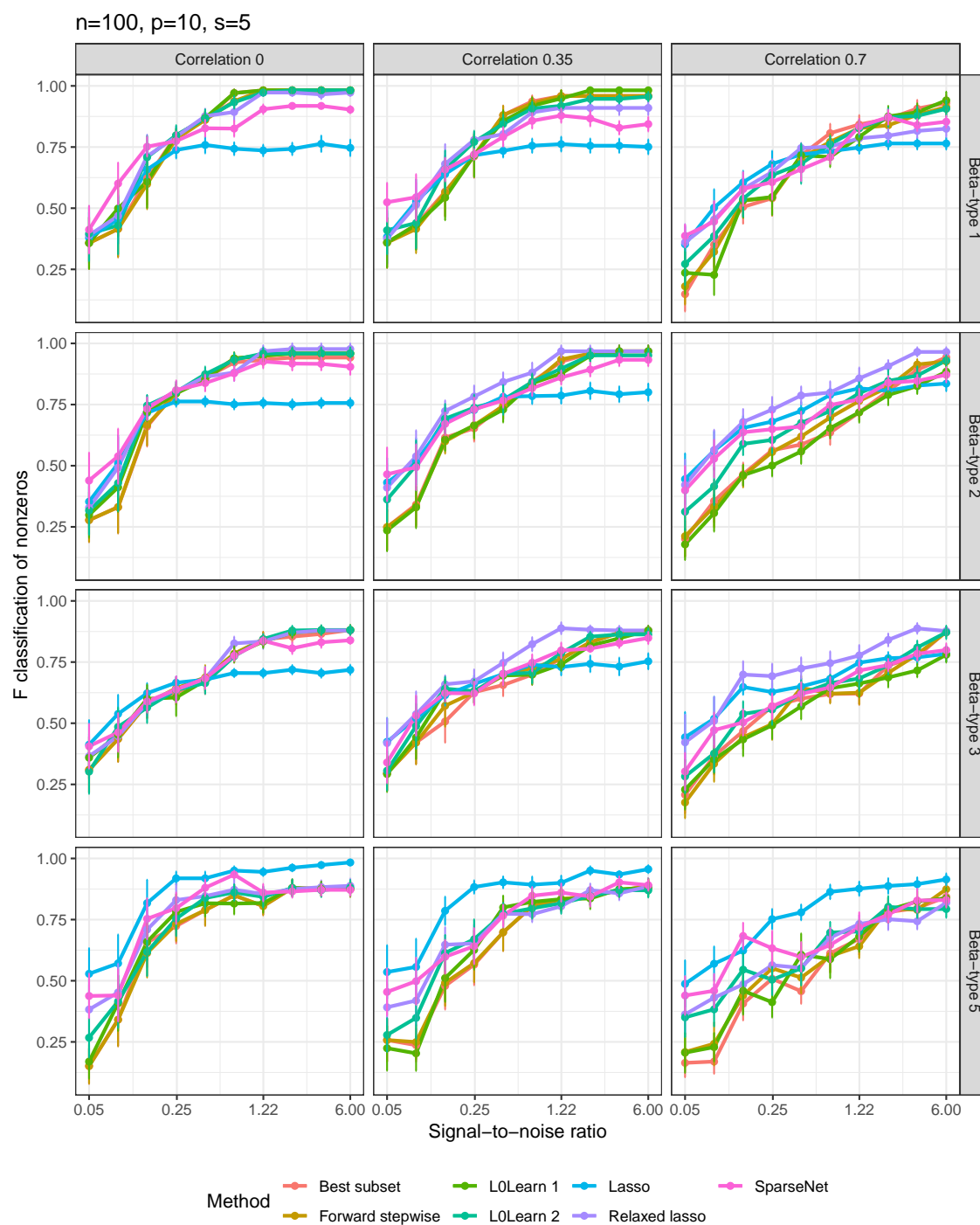
C.1.3 Proportion of variance explained



C.1.4 Number of nonzero coefficients

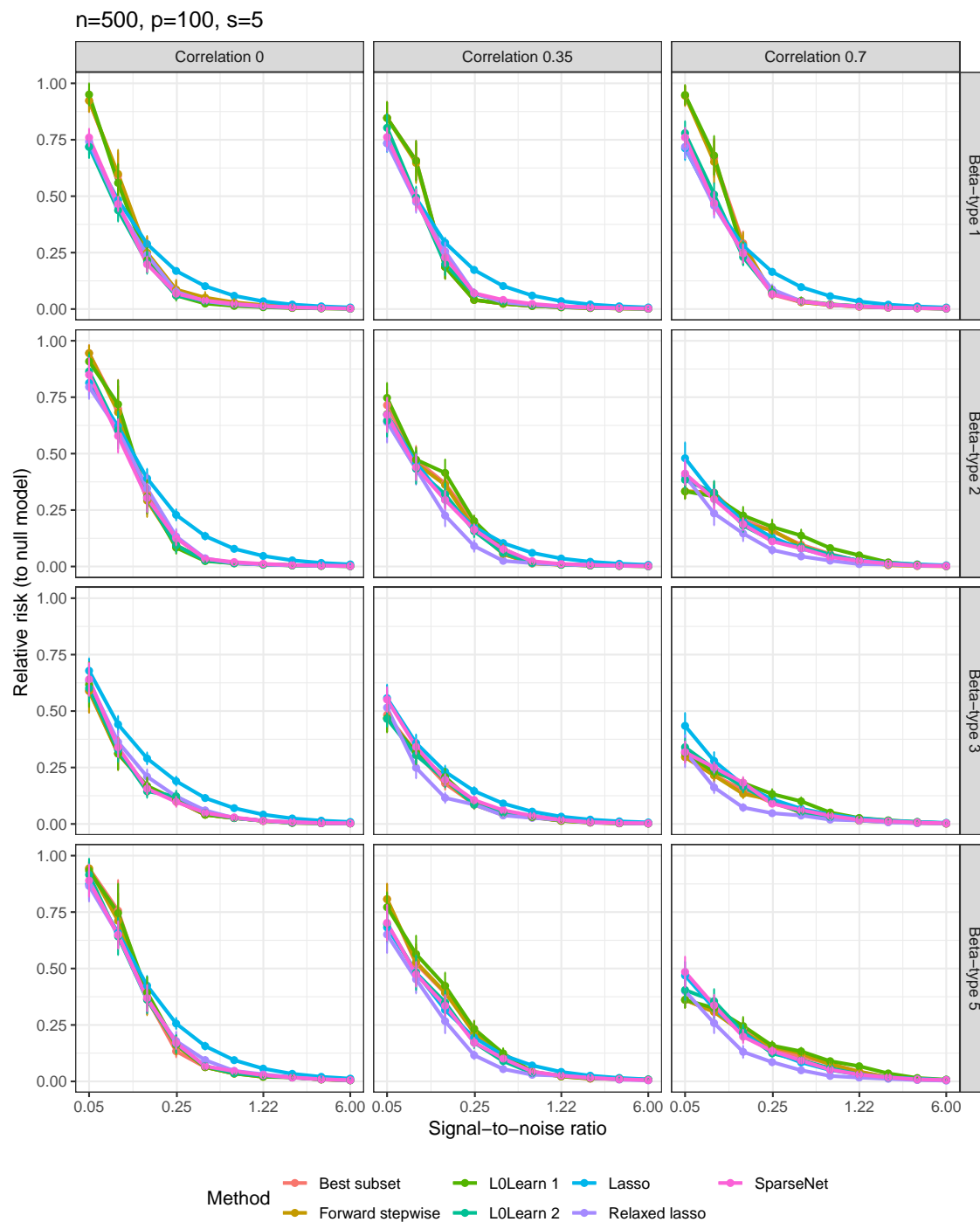


C.1.5 F-score

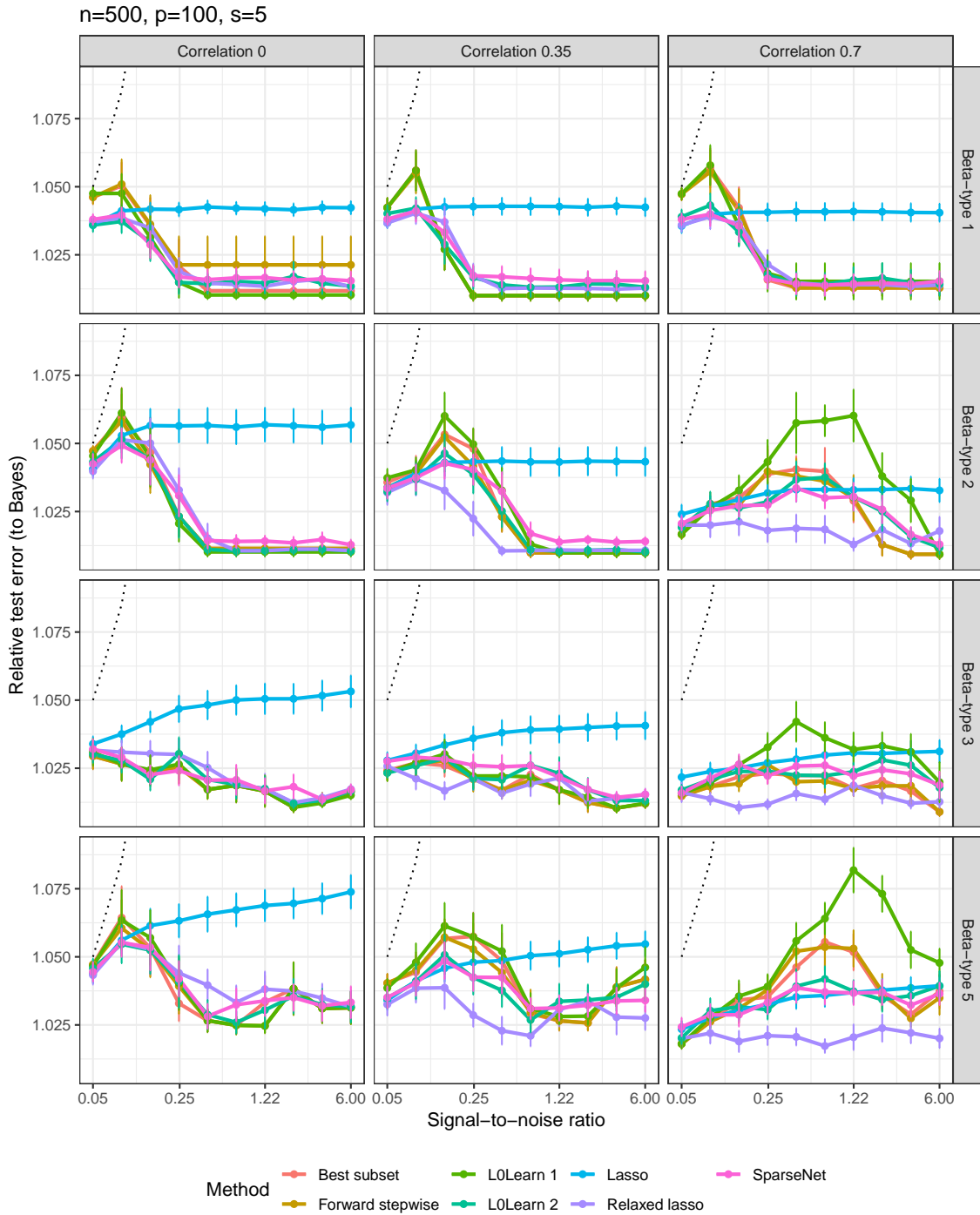


C.2 Medium setting: $n = 500, p = 100, s = 5$

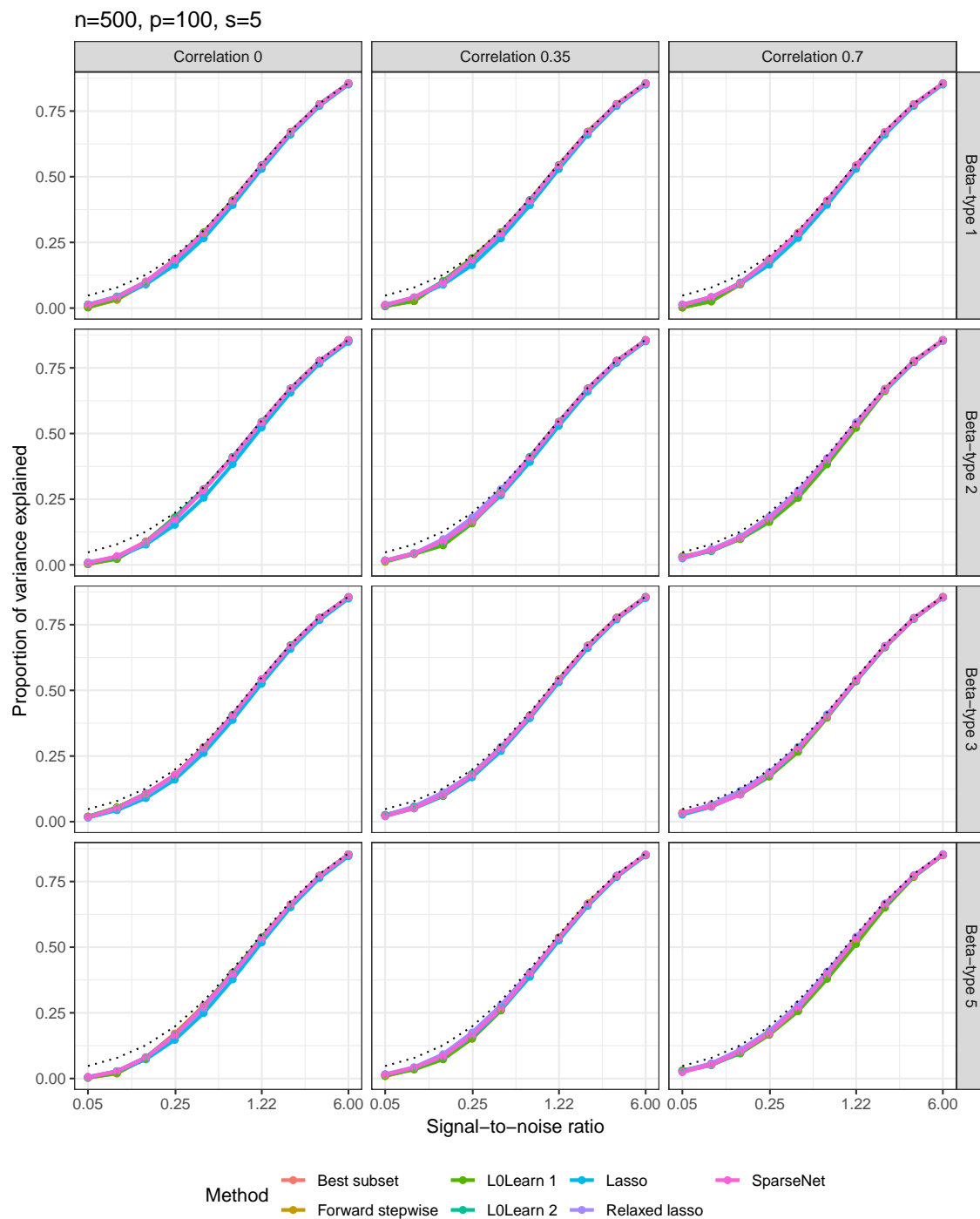
C.2.1 Relative risk (to null model)



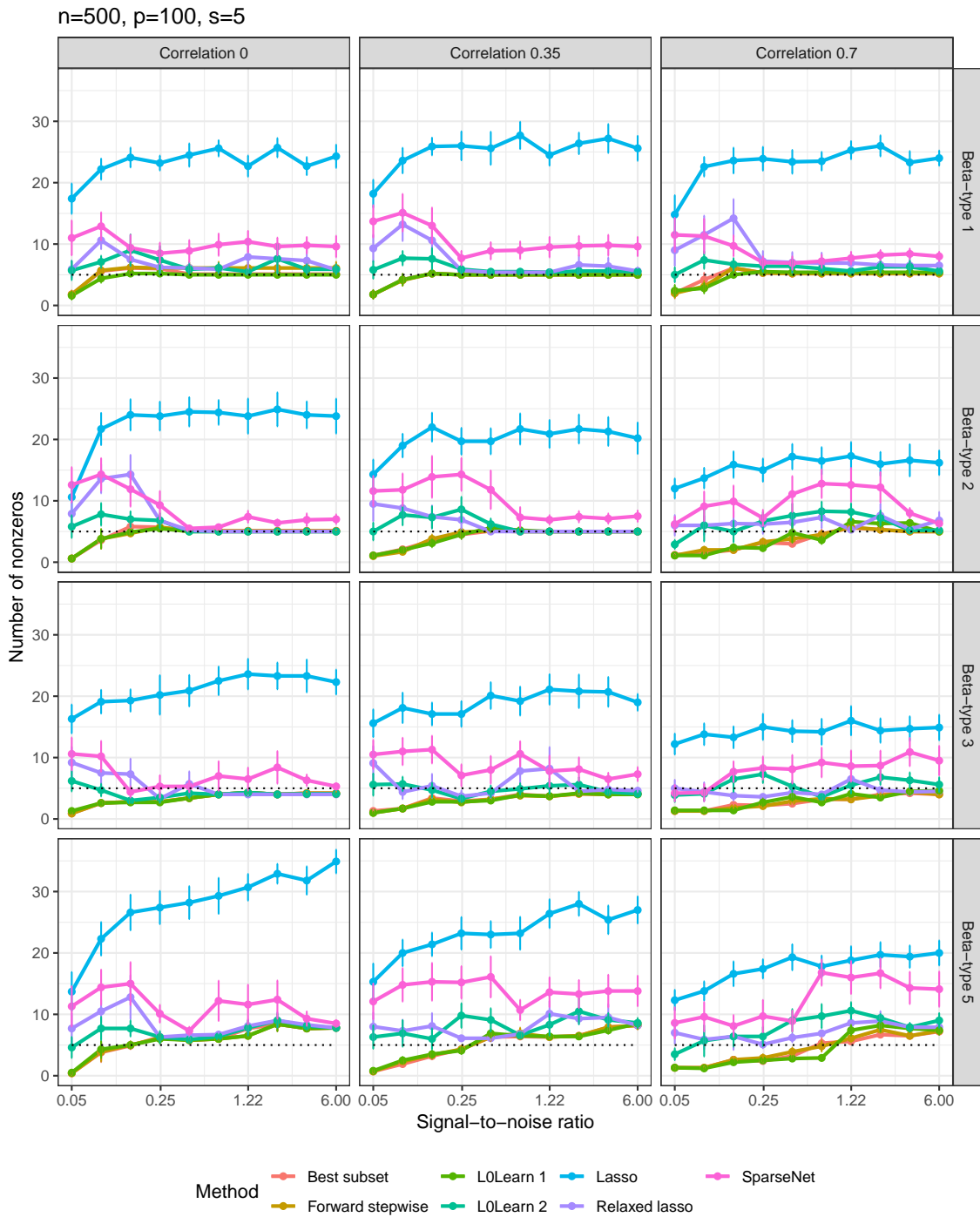
C.2.2 Relative test error (to Bayes)



C.2.3 Proportion of variance explained

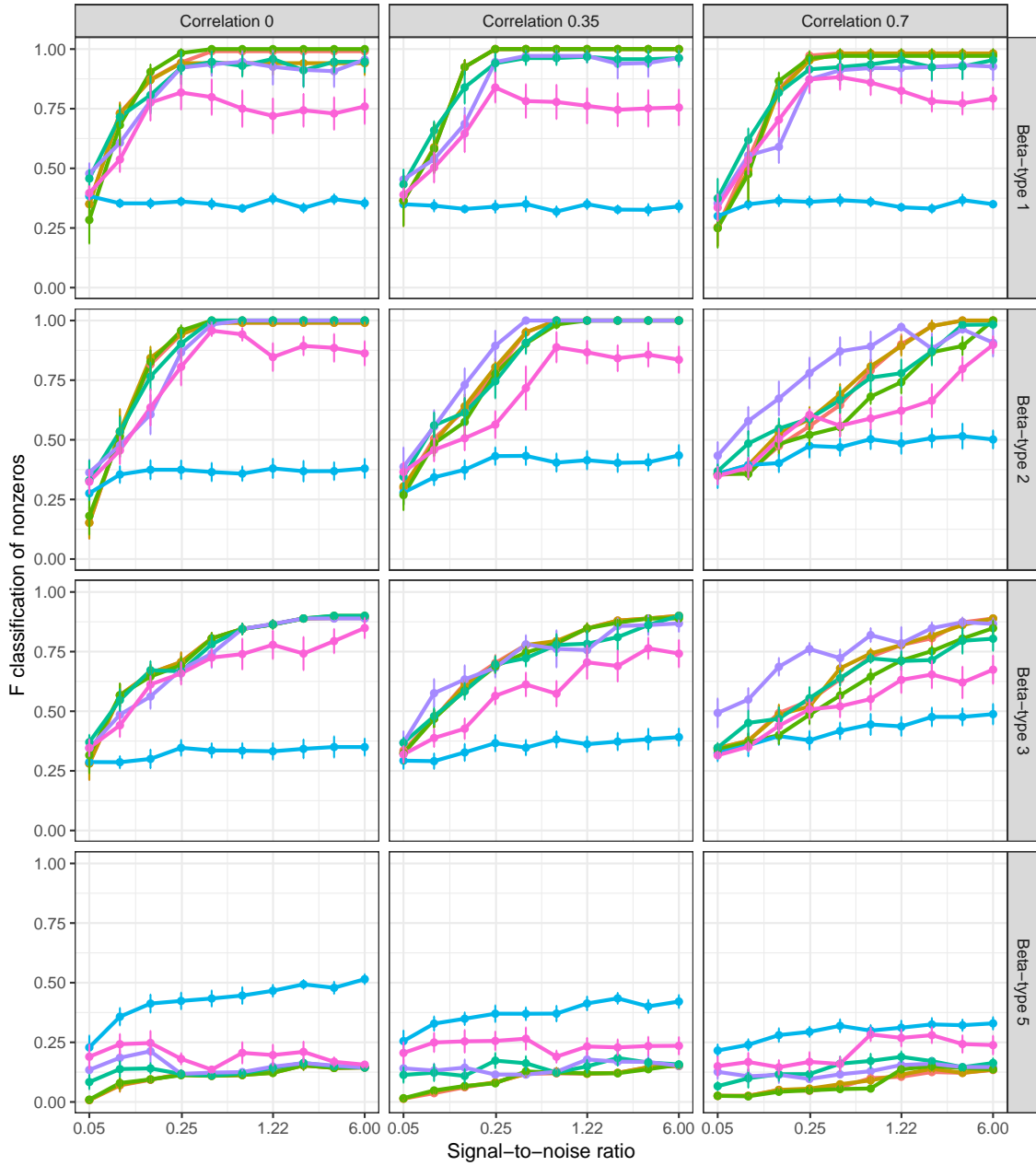


C.2.4 Number of nonzero coefficients



C.2.5 F-score

$n=500, p=100, s=5$

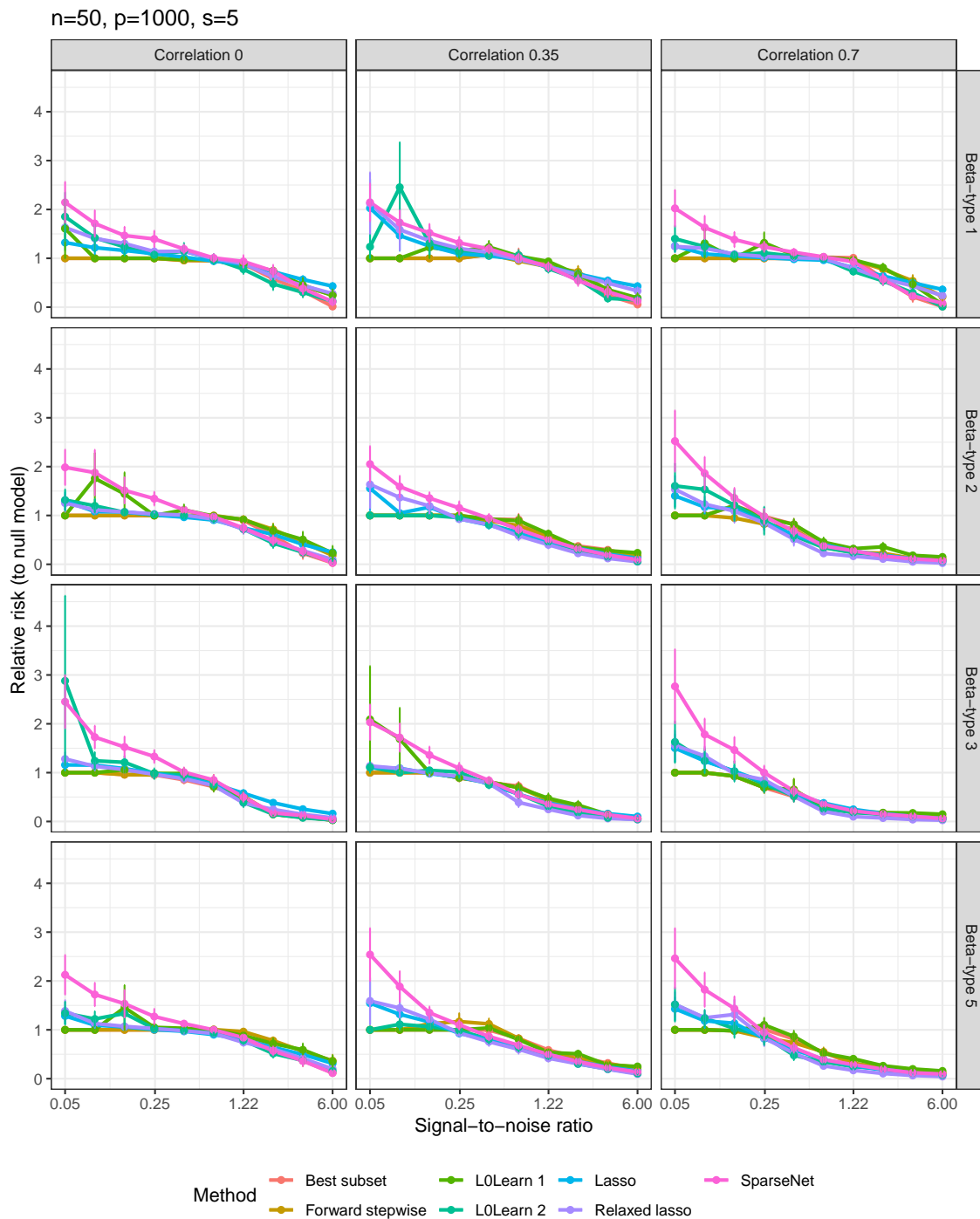


Method

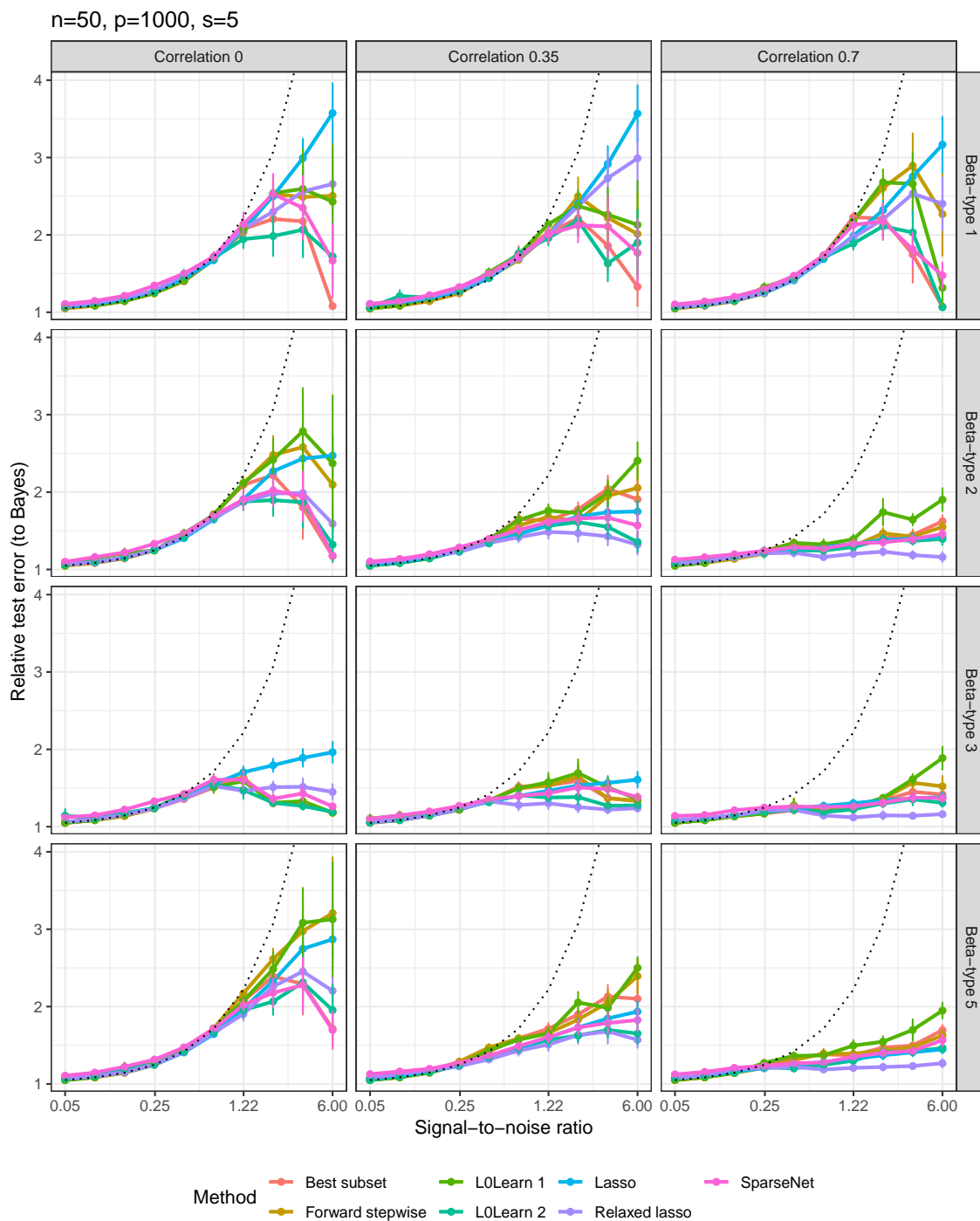
- Best subset
- L0Learn 1
- Lasso
- SparseNet
- Forward stepwise
- L0Learn 2
- Relaxed lasso

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

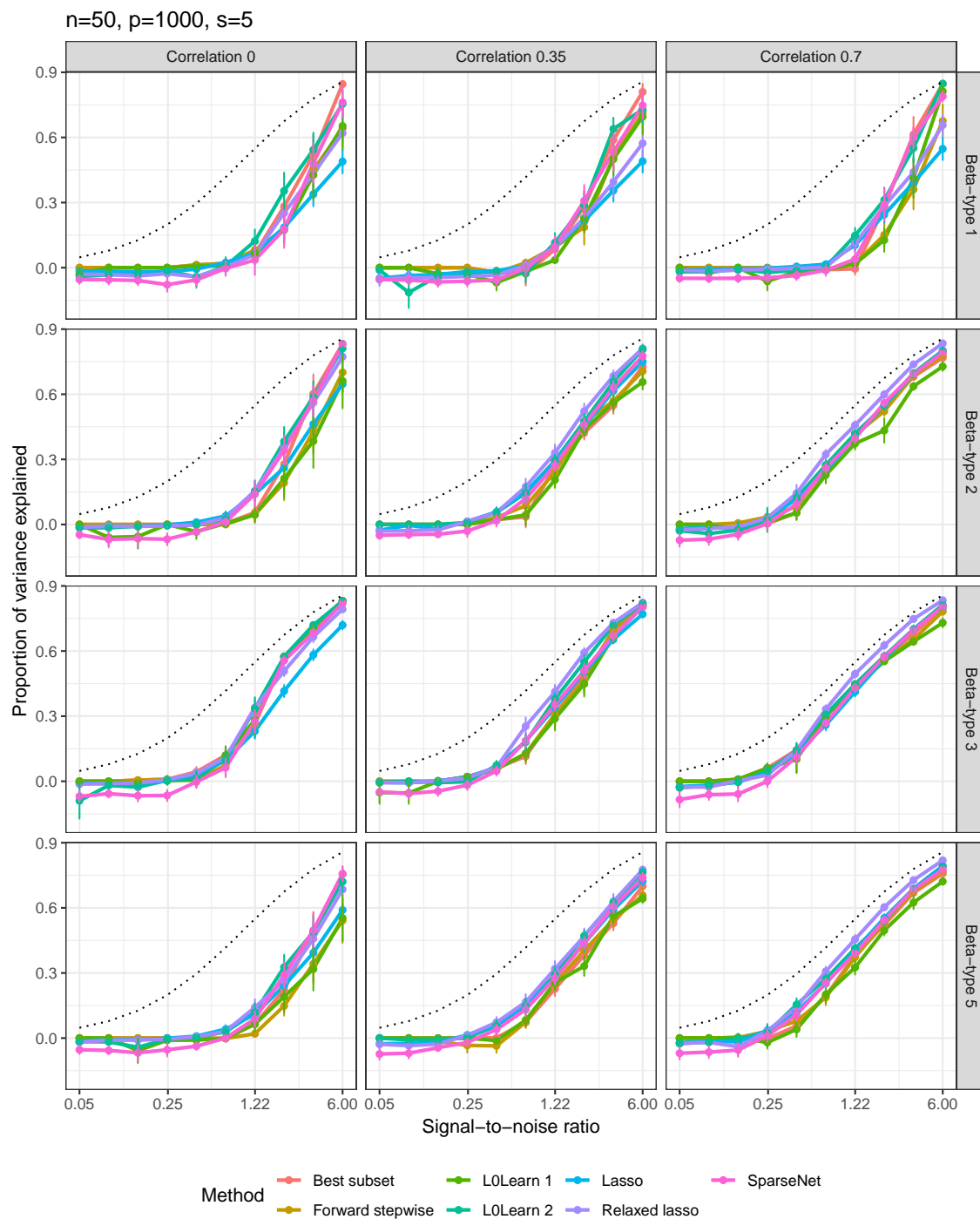
C.3.1 Relative risk (to null model)



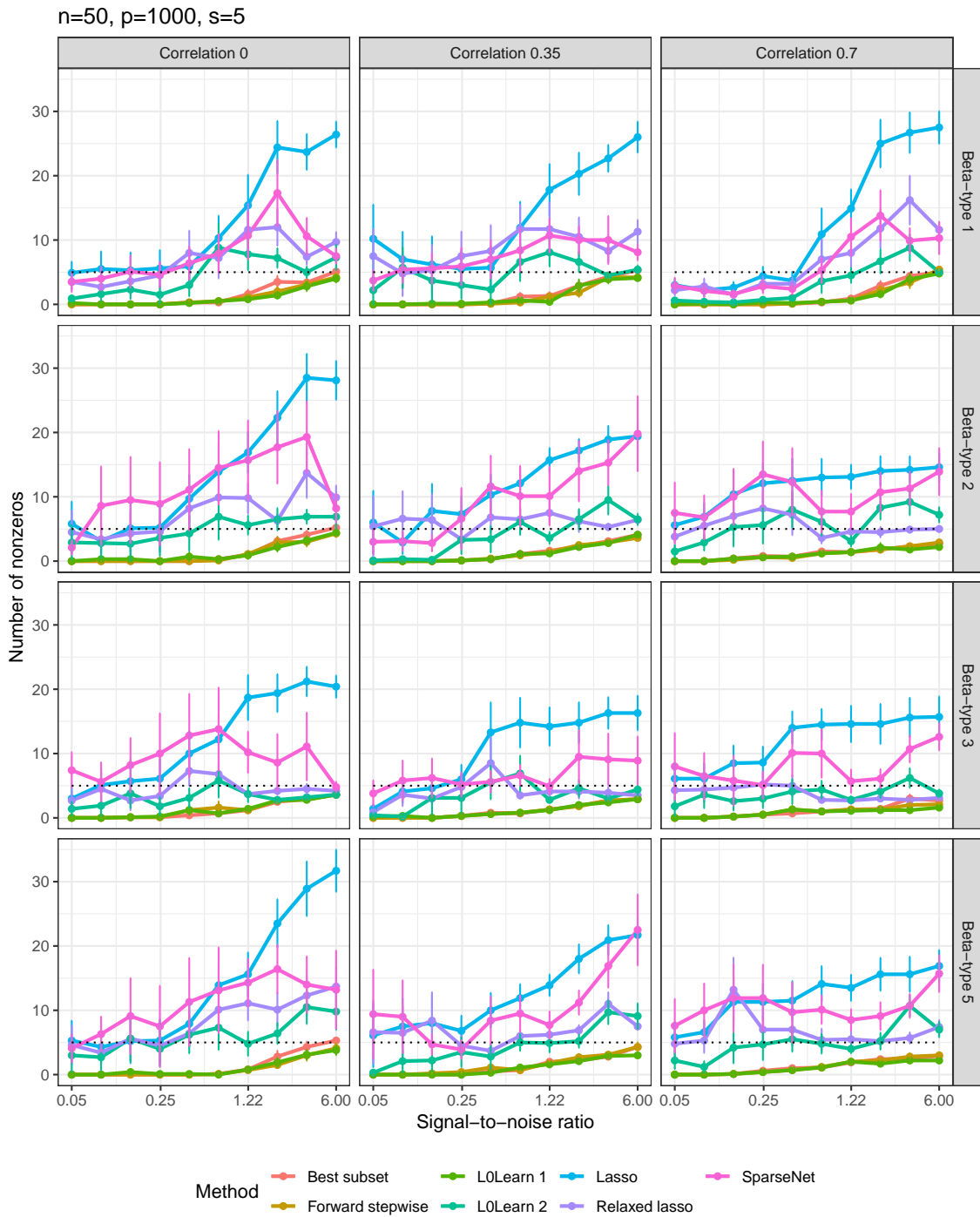
C.3.2 Relative test error (to Bayes)



C.3.3 Proportion of variance explained

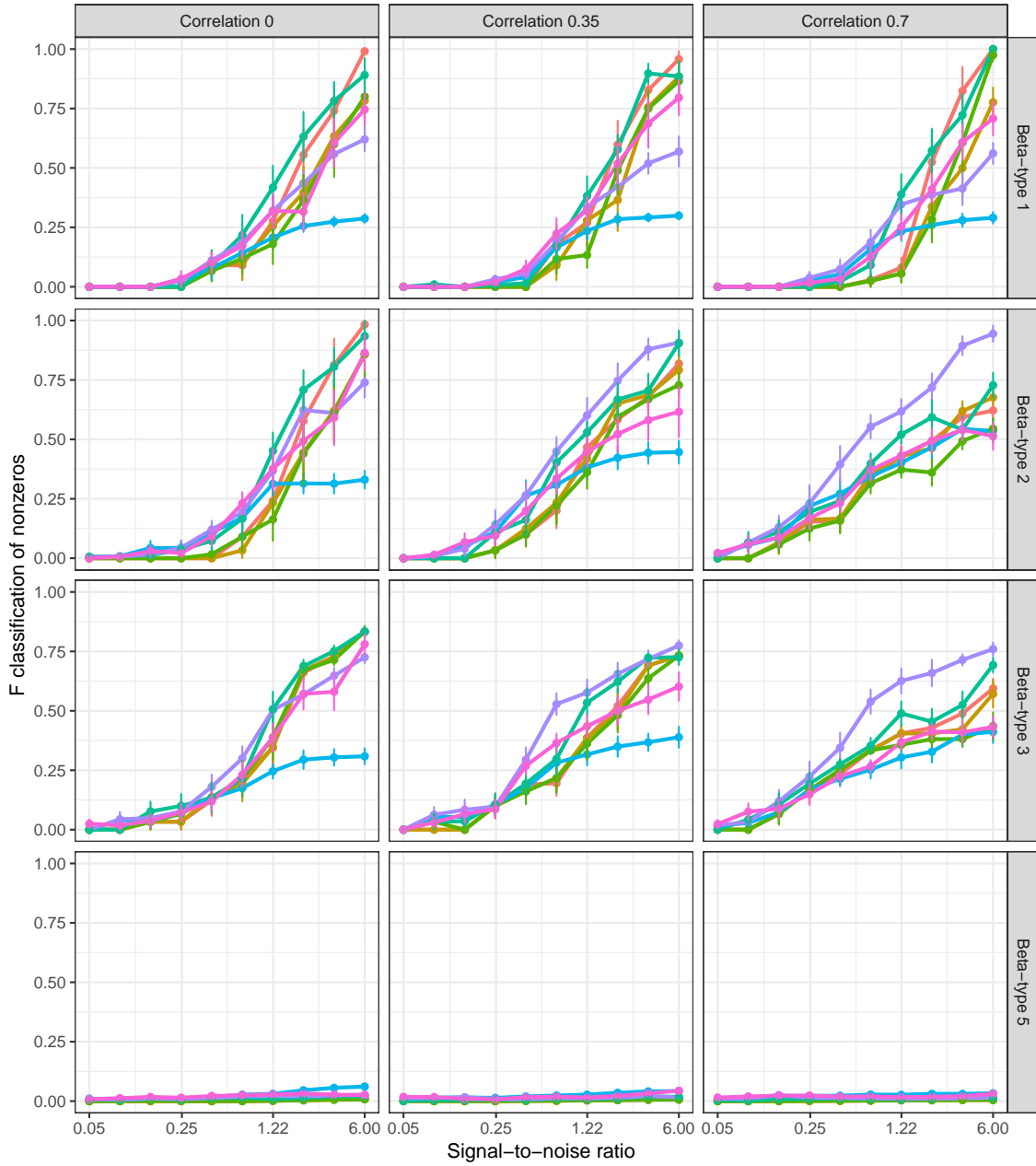


C.3.4 Number of nonzero coefficients



C.3.5 F-score

$n=50, p=1000, s=5$

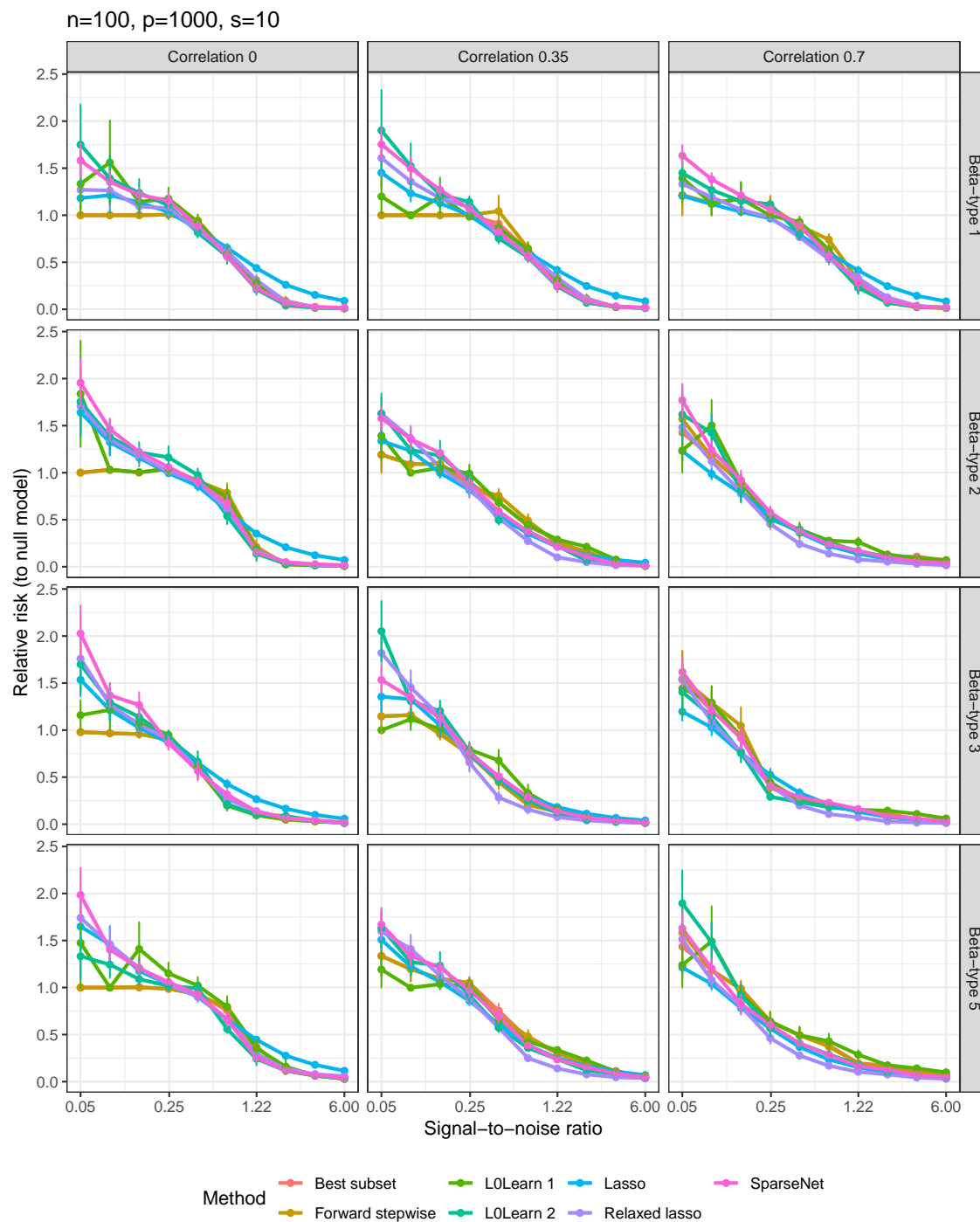


Method

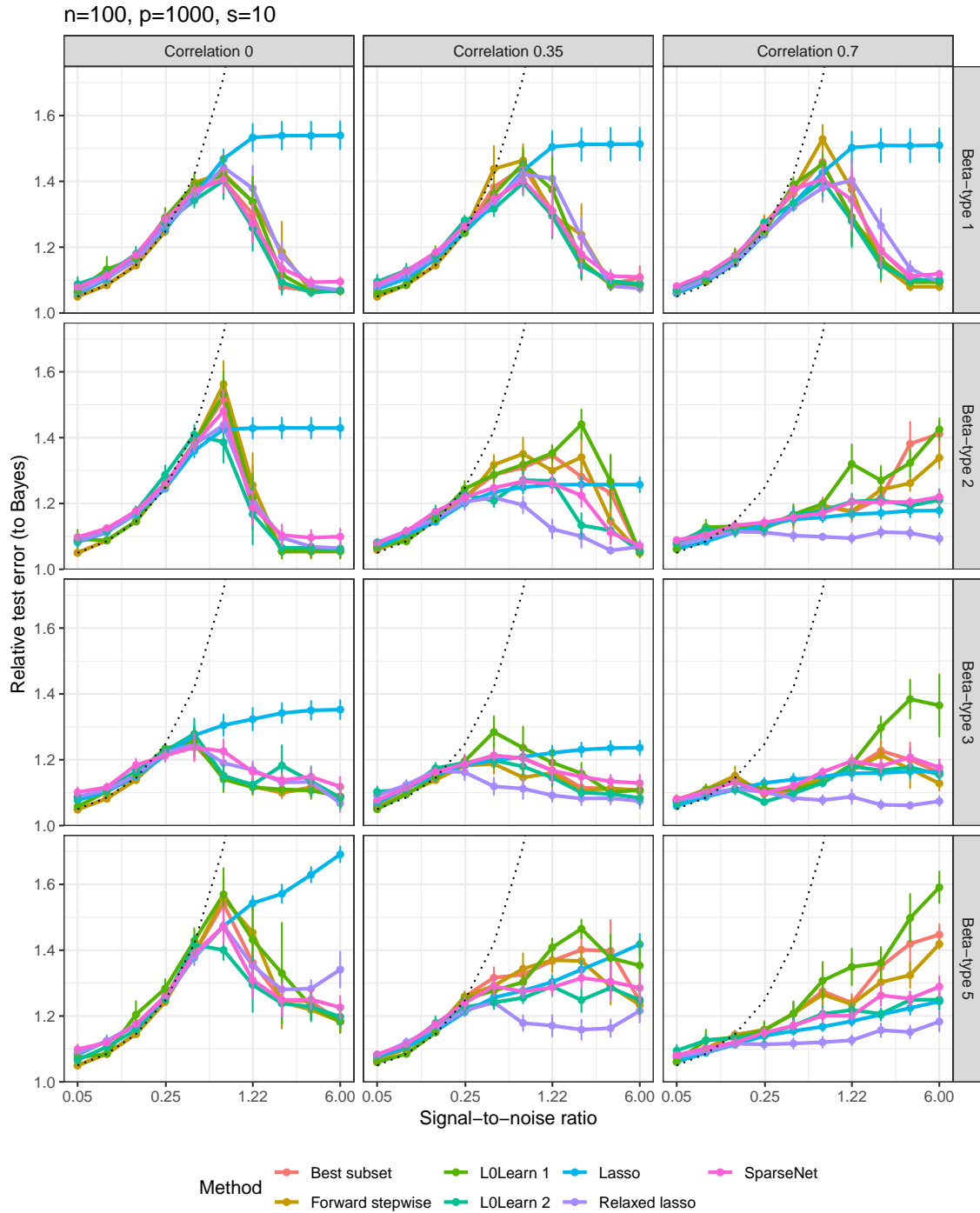
- Best subset
- L0Learn 1
- Lasso
- SparseNet
- Forward stepwise
- L0Learn 2
- Relaxed lasso

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

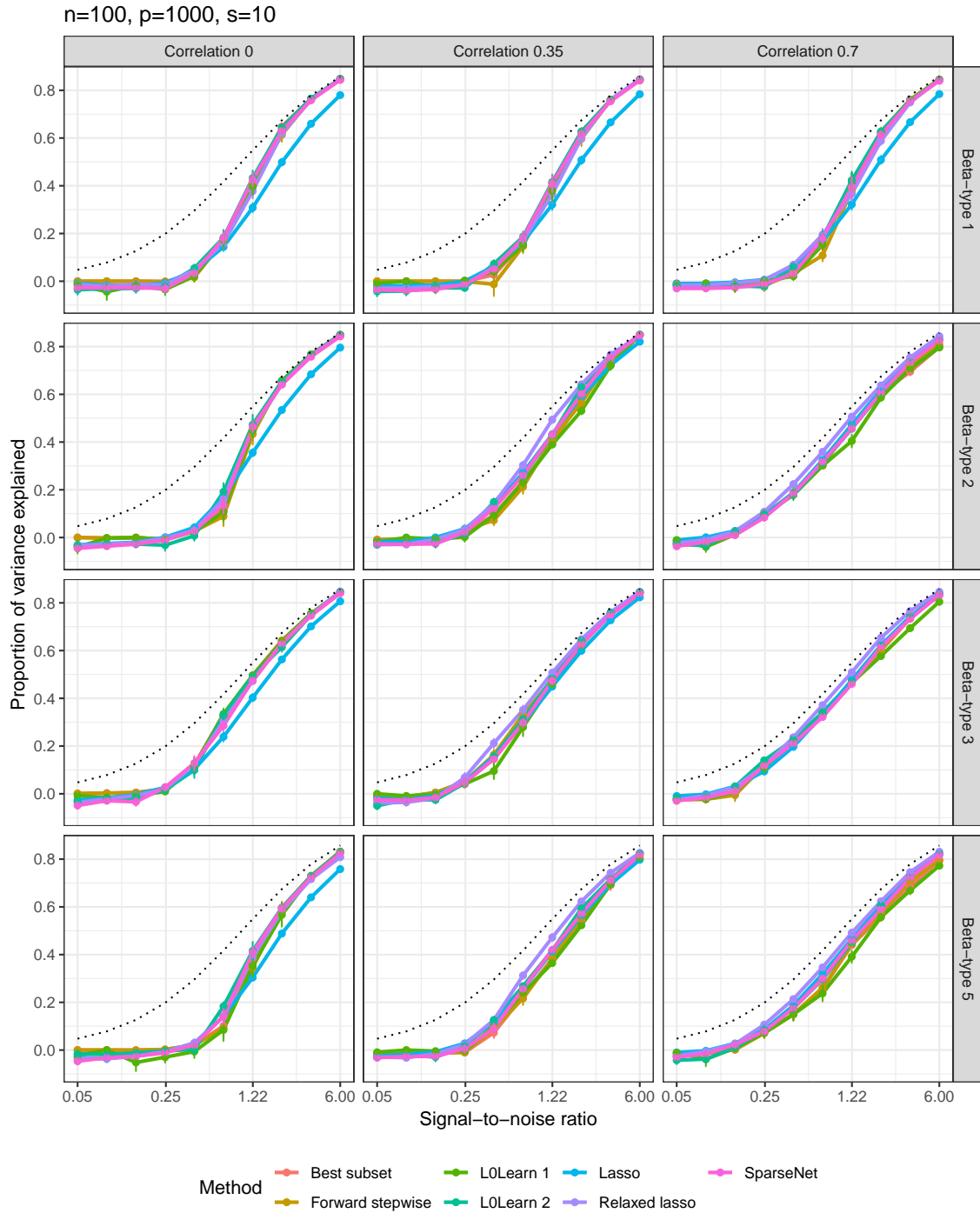
C.4.1 Relative risk (to null model)



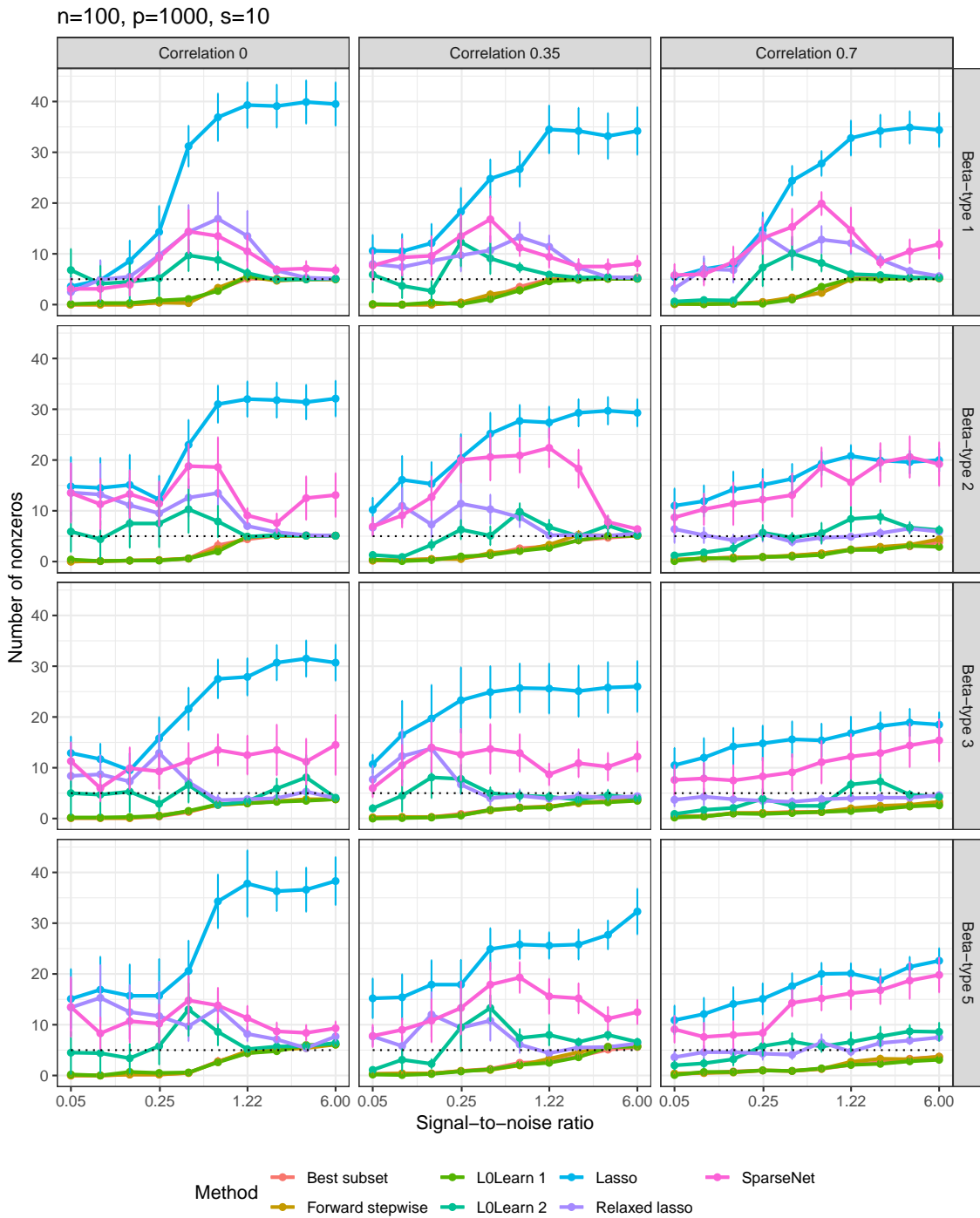
C.4.2 Relative test error (to Bayes)



C.4.3 Proportion of variance explained

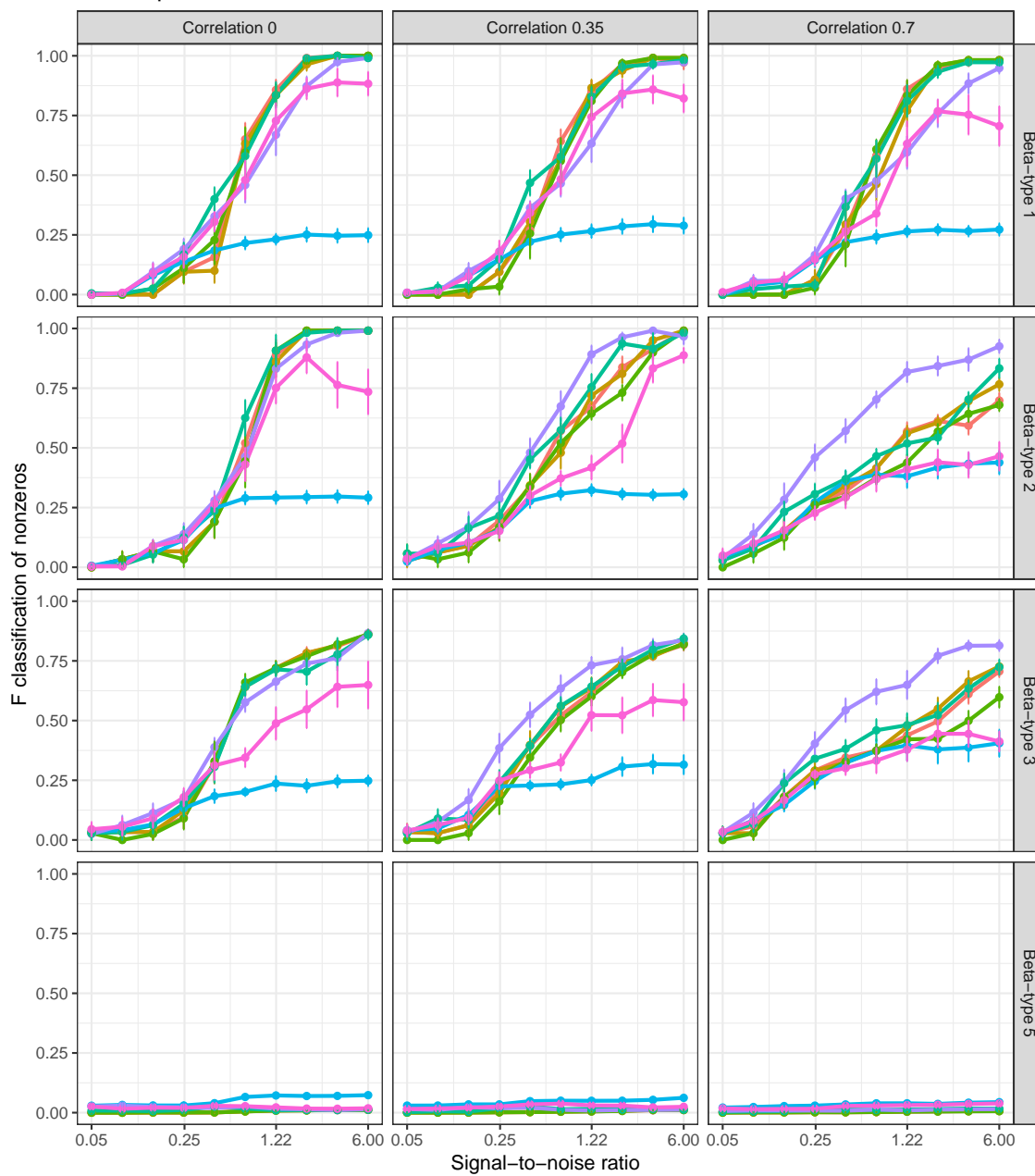


C.4.4 Number of nonzero coefficients



C.4.5 F-score

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



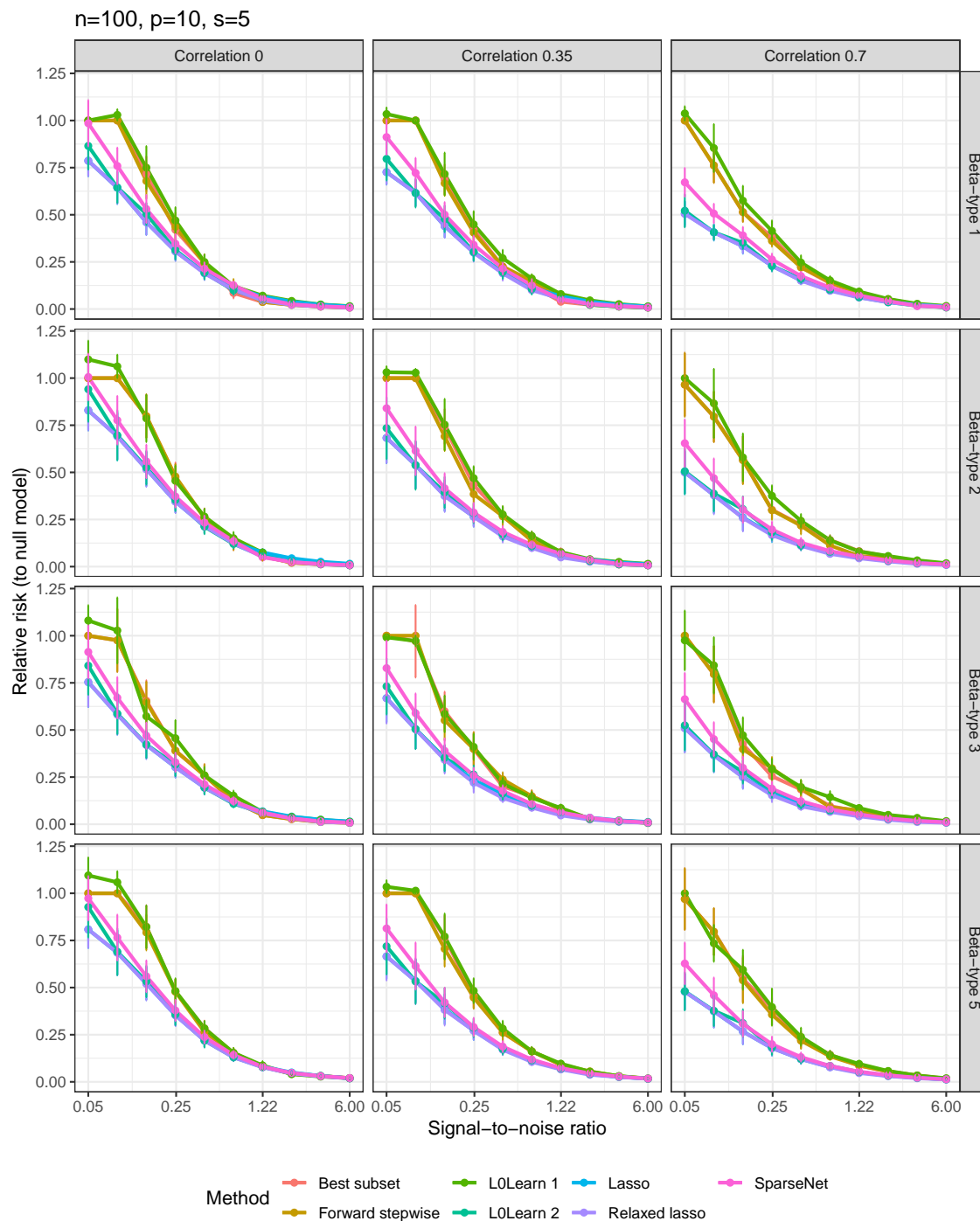
Method

- Best subset
- L0Learn 1
- Lasso
- SparseNet
- Forward stepwise
- L0Learn 2
- Relaxed lasso

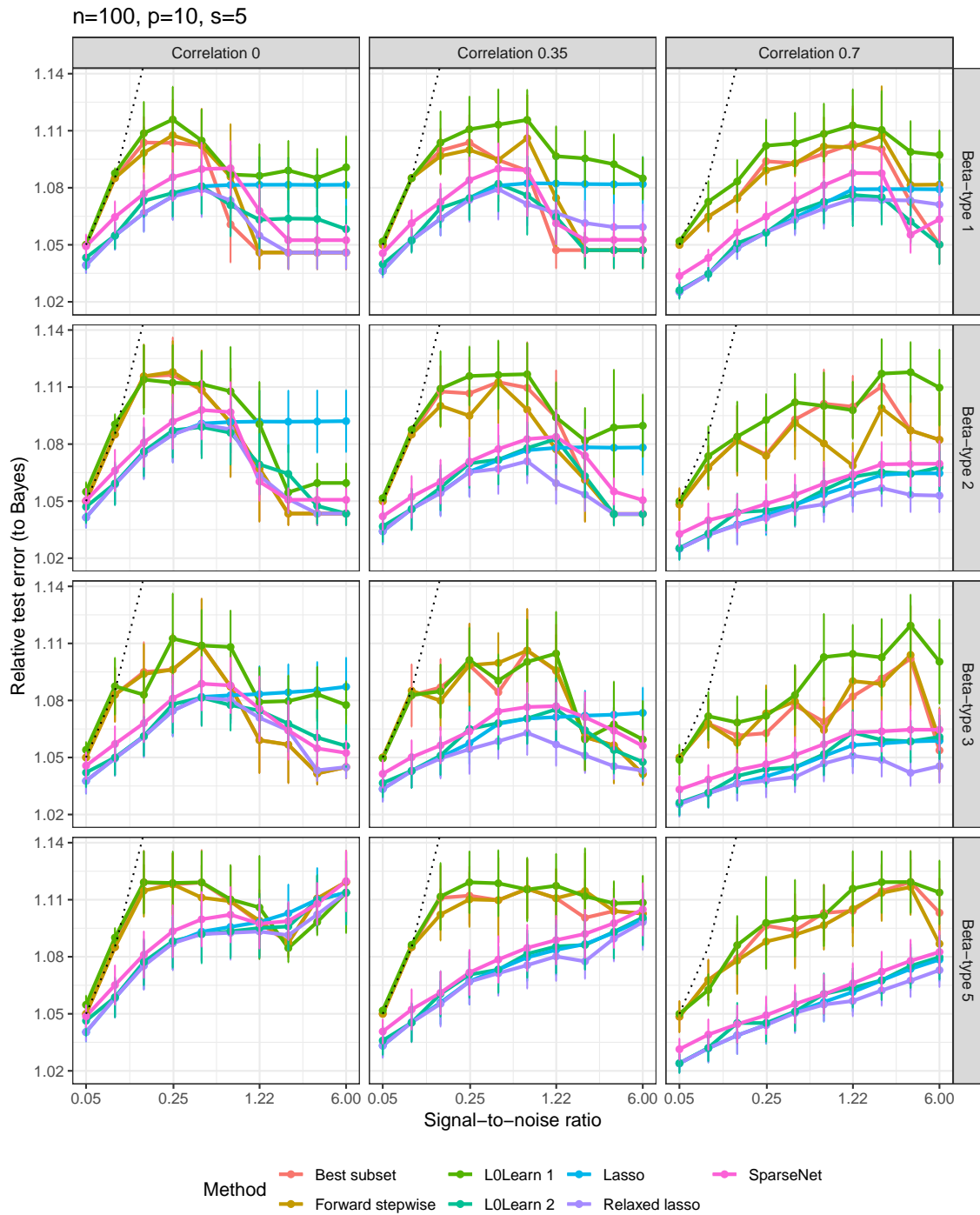
D Oracle tuning: more methods

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

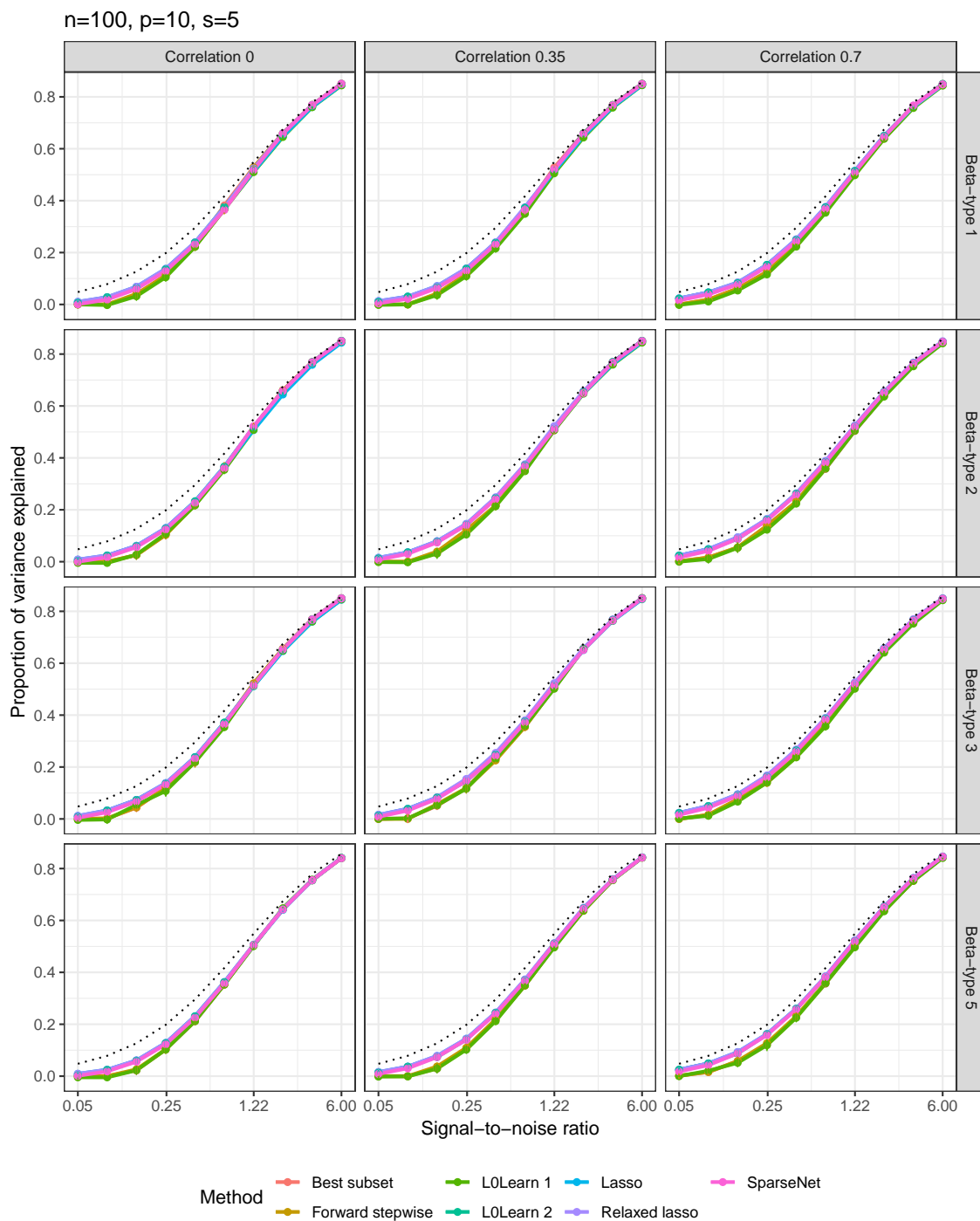
D.1.1 Relative risk (to null model)



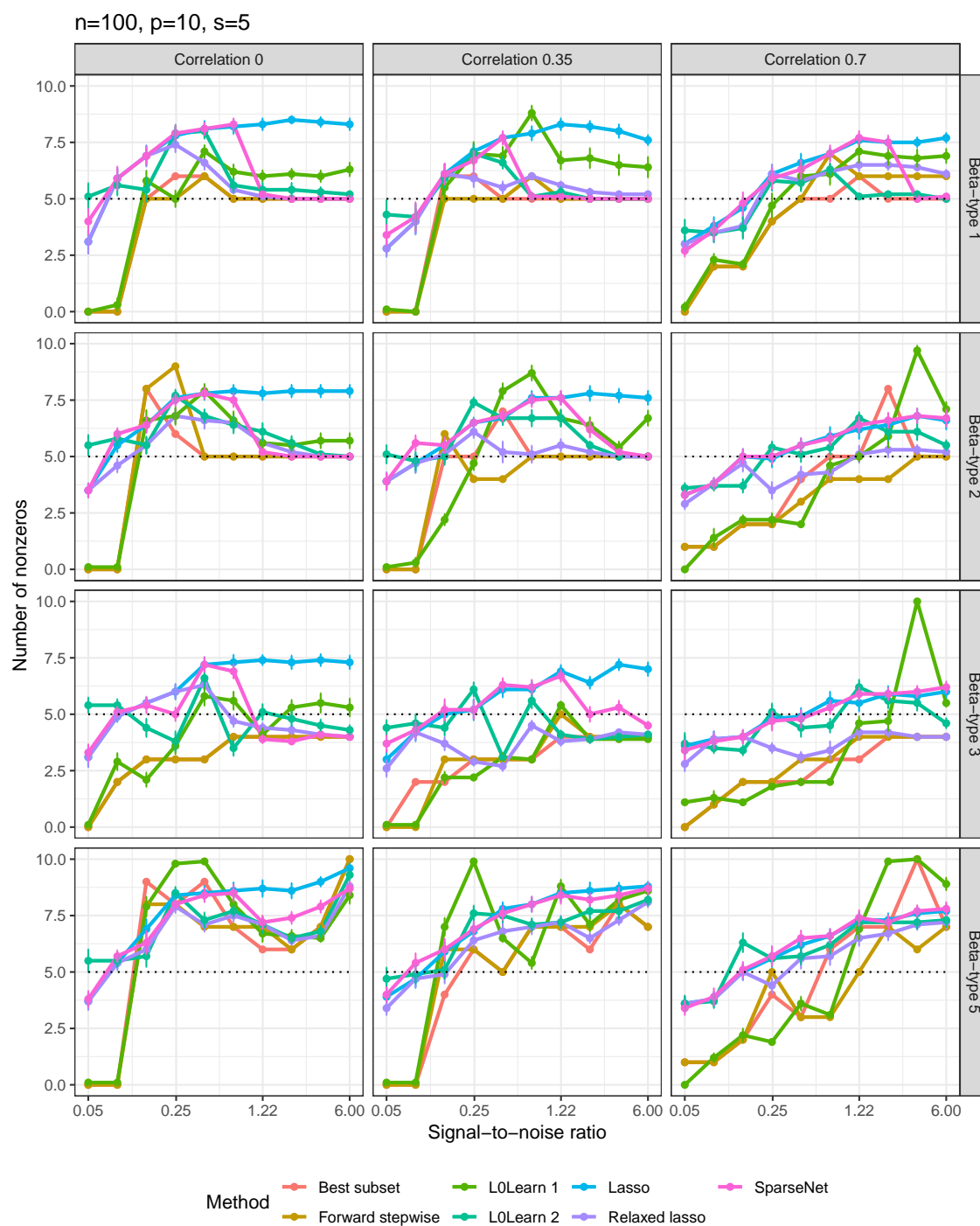
D.1.2 Relative test error (to Bayes)



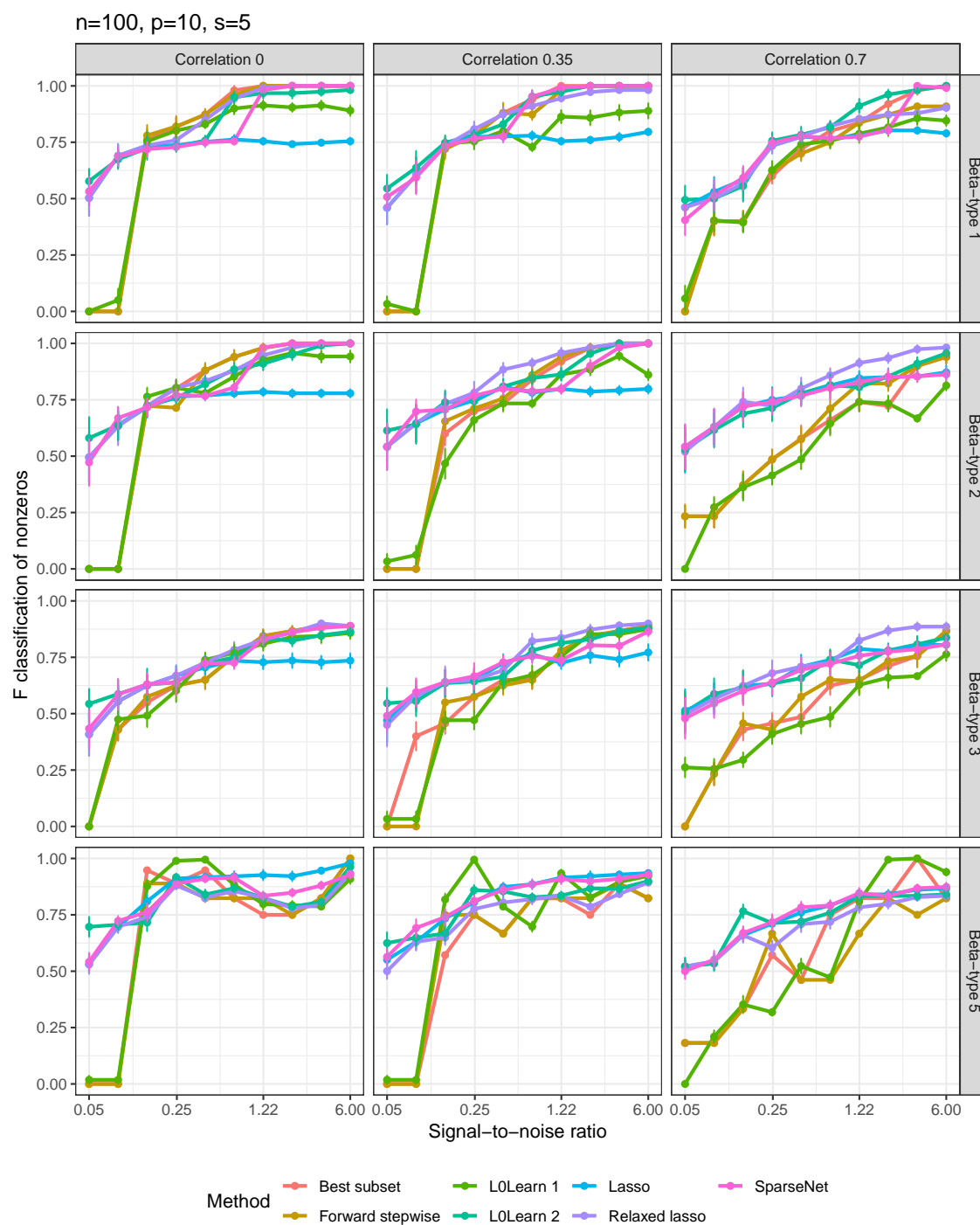
D.1.3 Proportion of variance explained



D.1.4 Number of nonzero coefficients

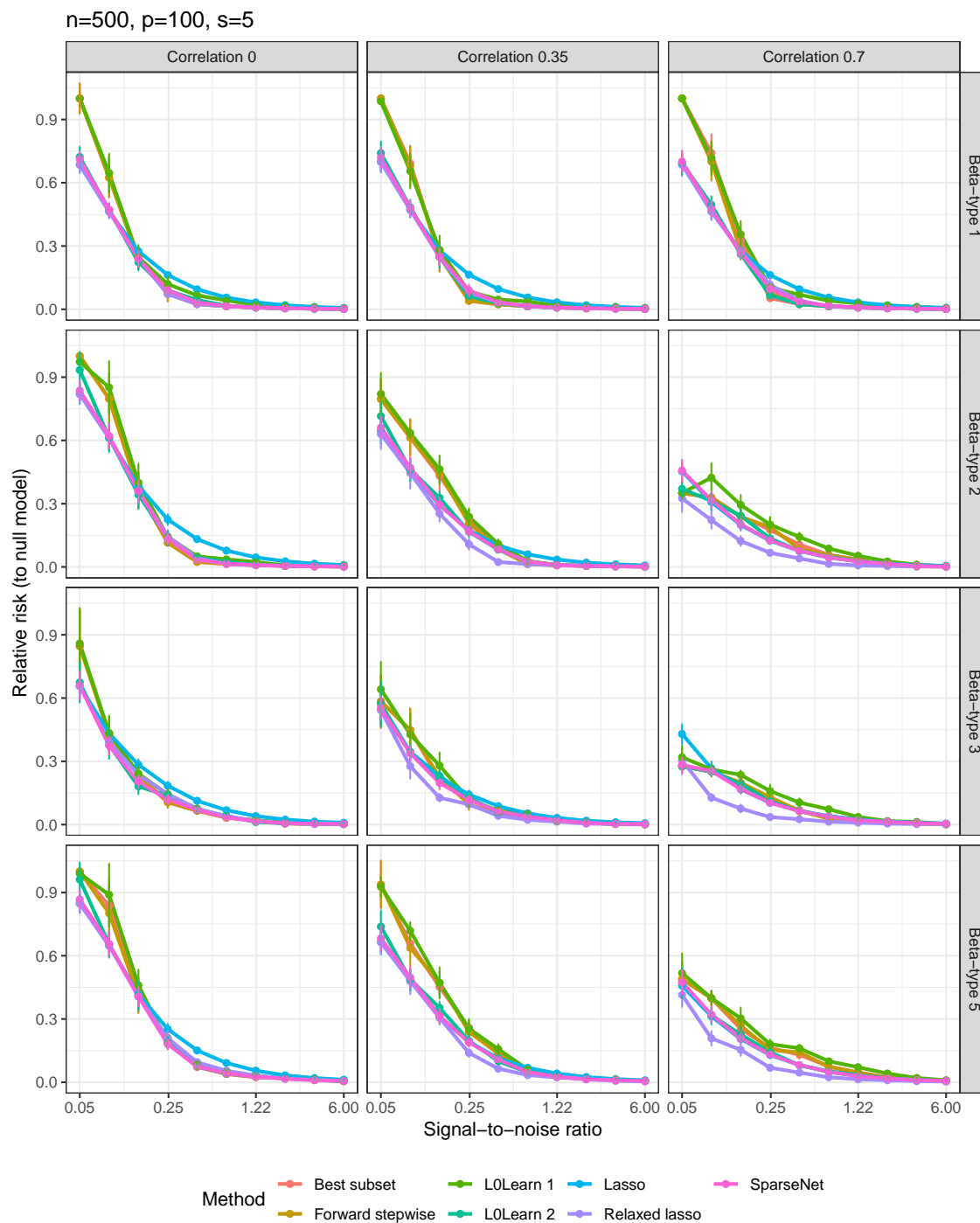


D.1.5 F-score

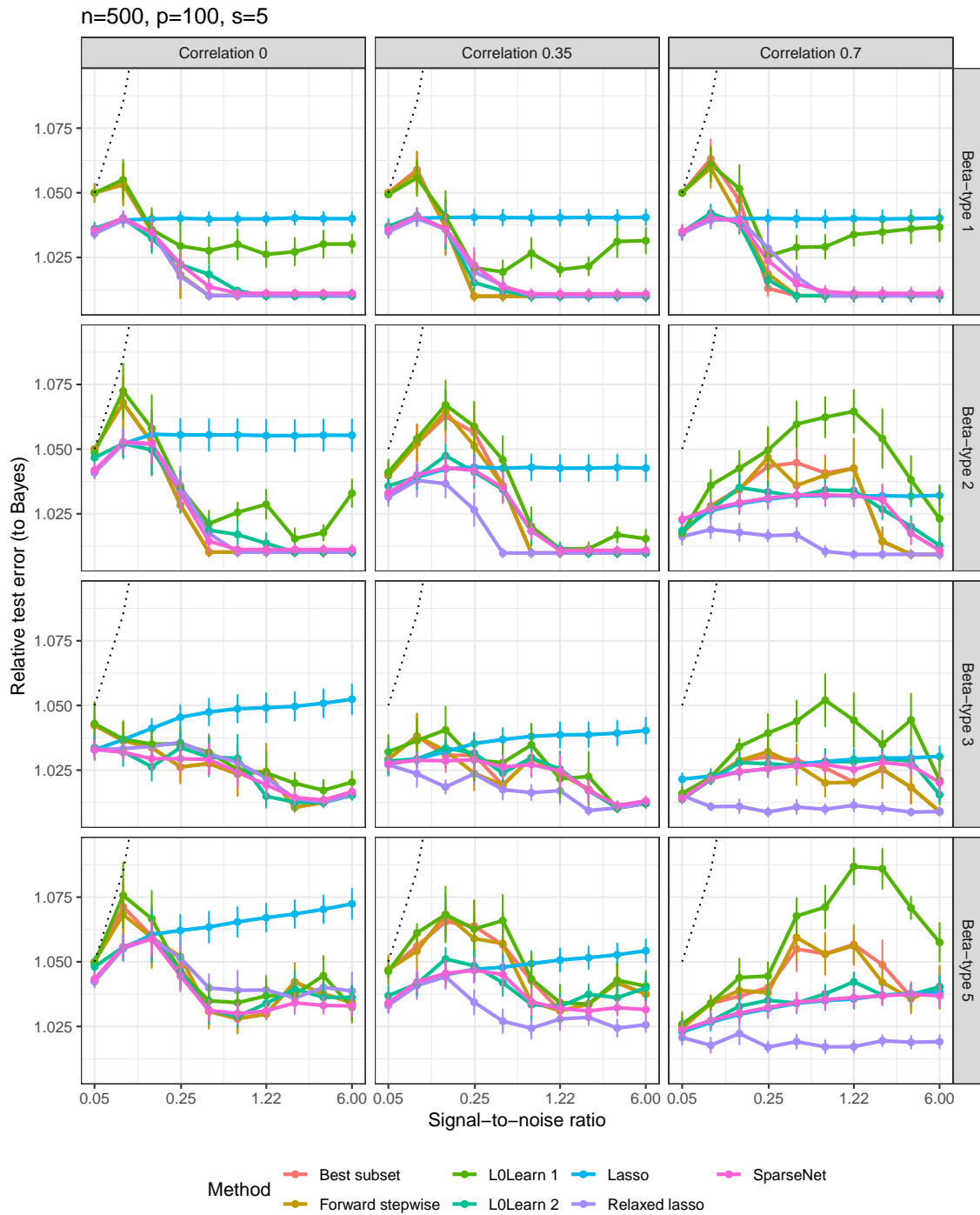


D.2 Medium setting: $n = 500, p = 100, s = 5$

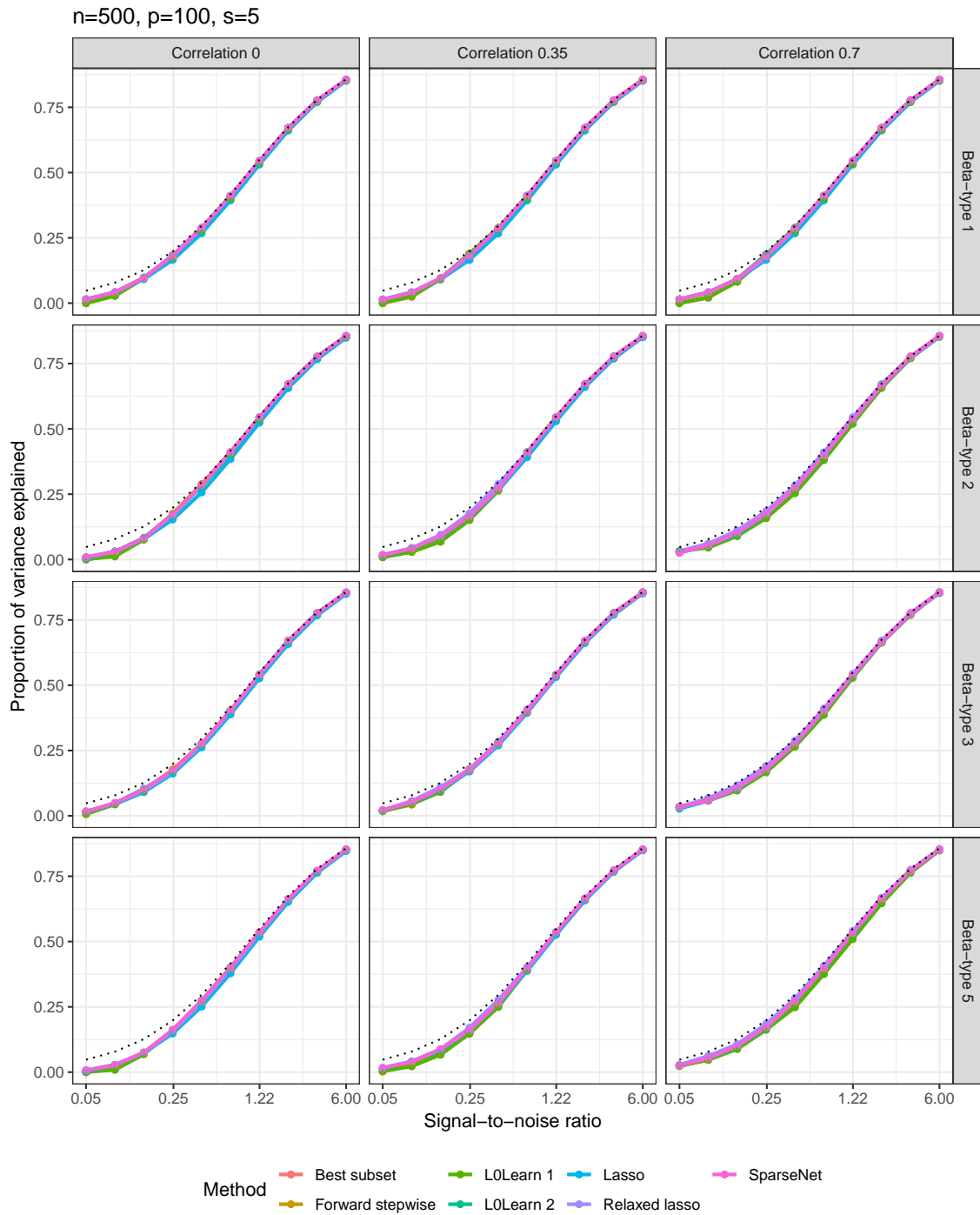
D.2.1 Relative risk (to null model)



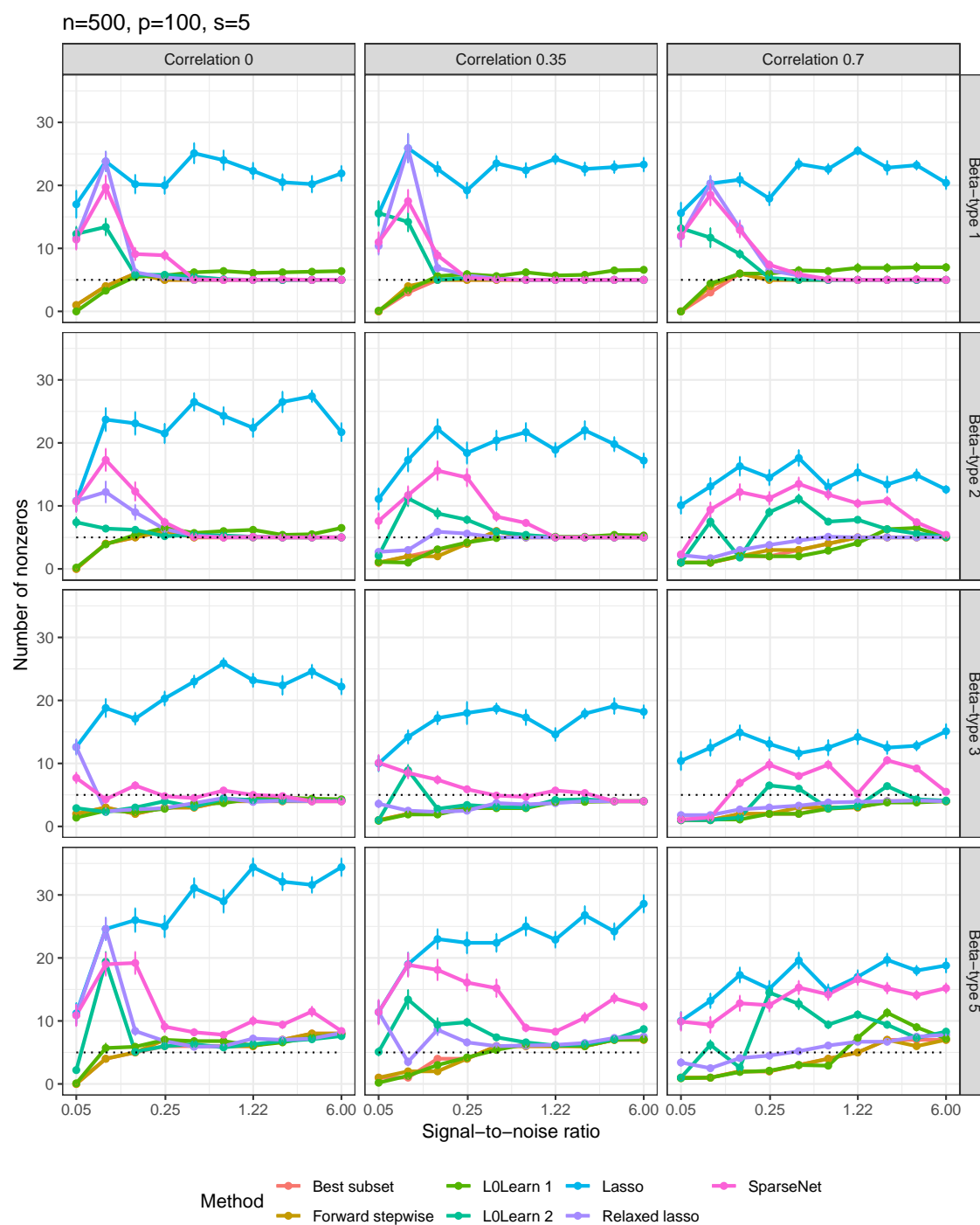
D.2.2 Relative test error (to Bayes)



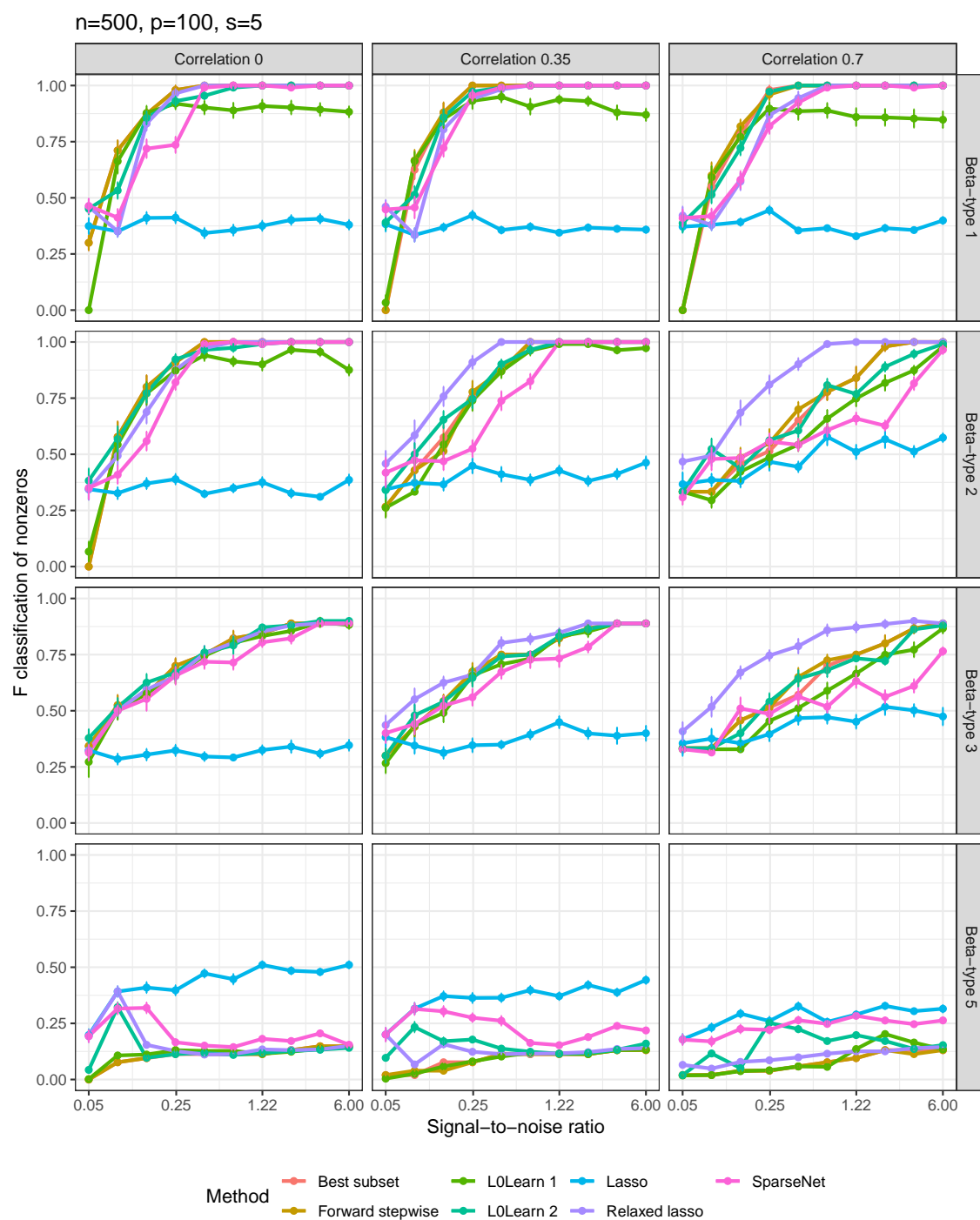
D.2.3 Proportion of variance explained



D.2.4 Number of nonzero coefficients

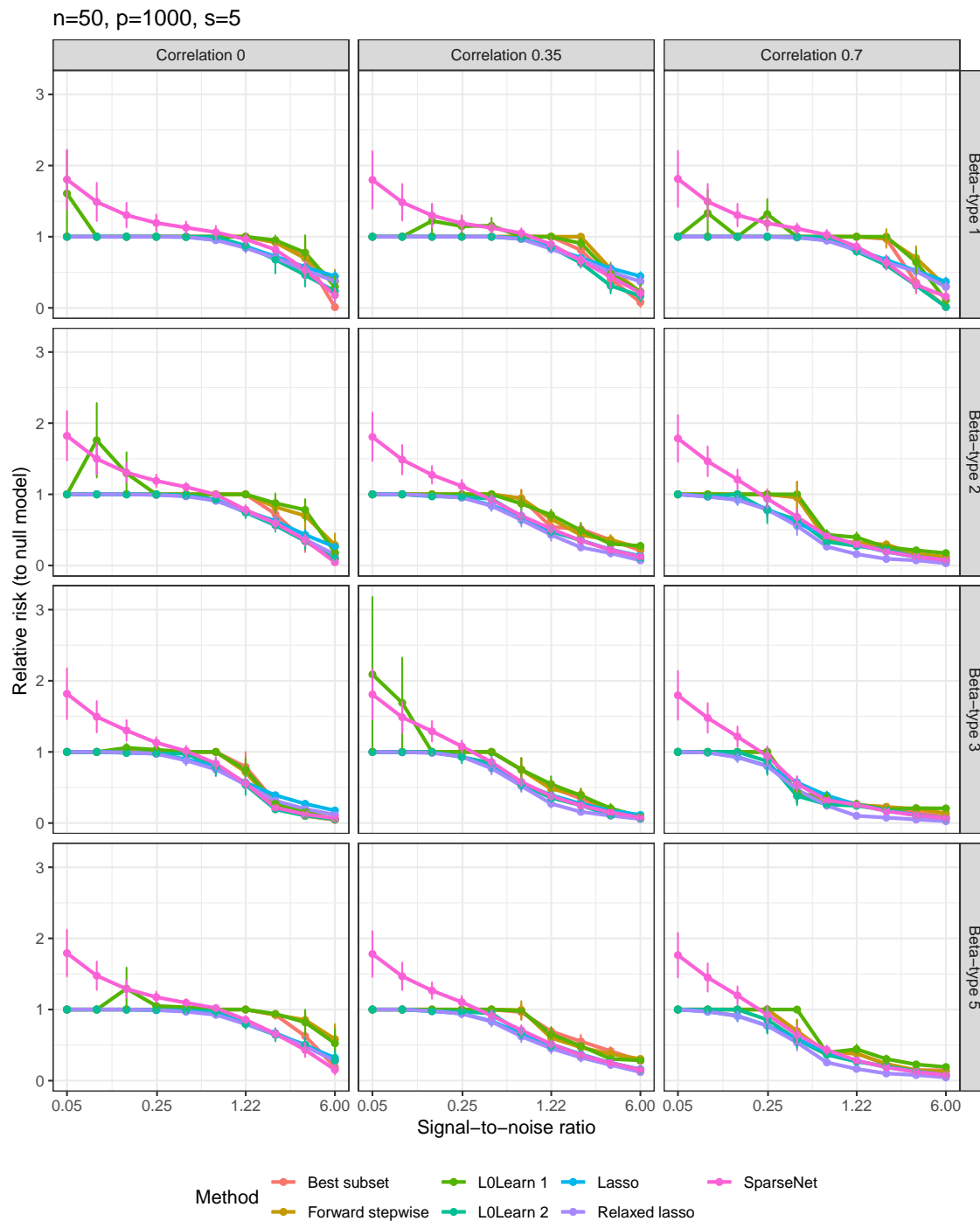


D.2.5 F-score

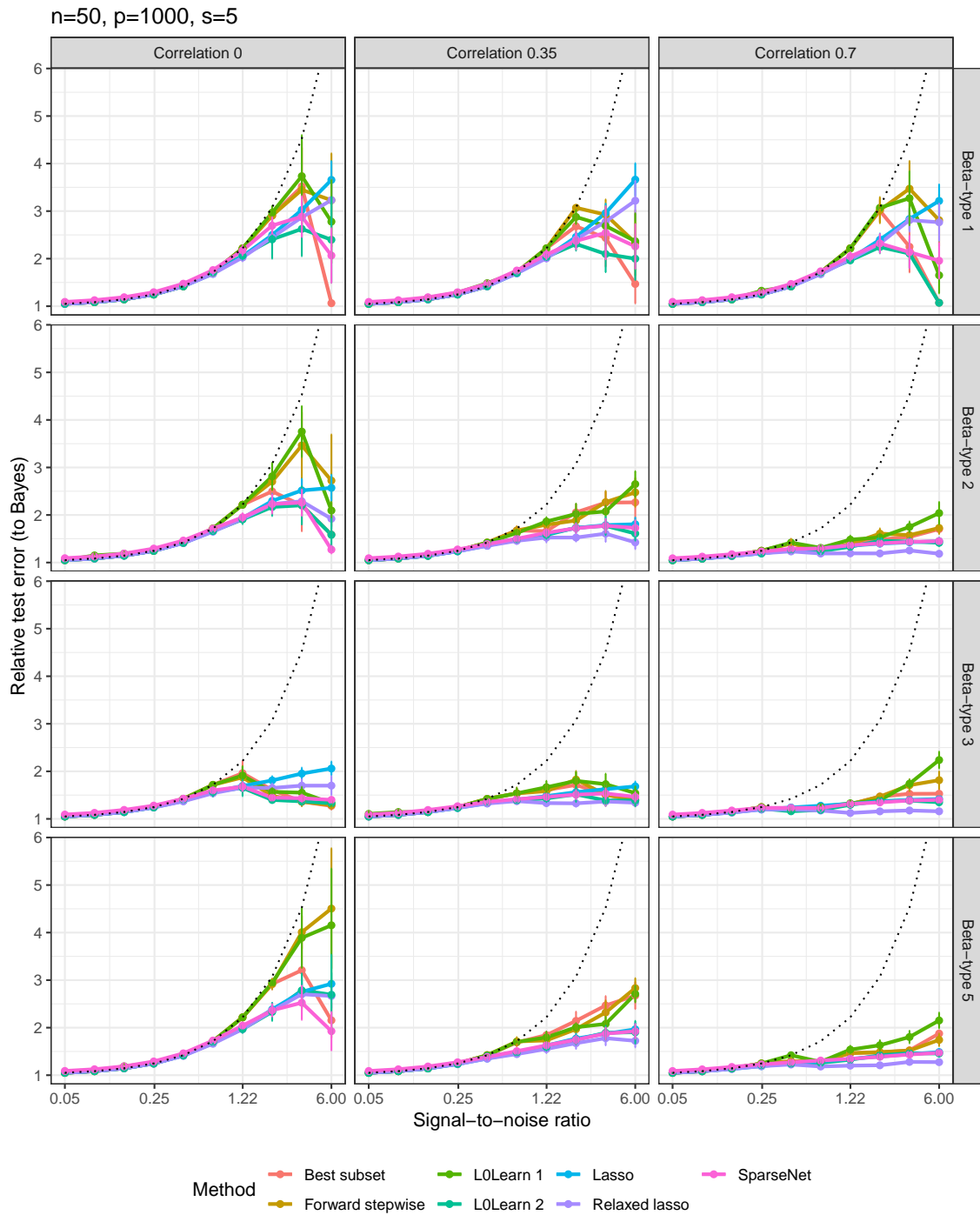


D.3 High-5 setting: $n = 50, p = 1000, s = 5$

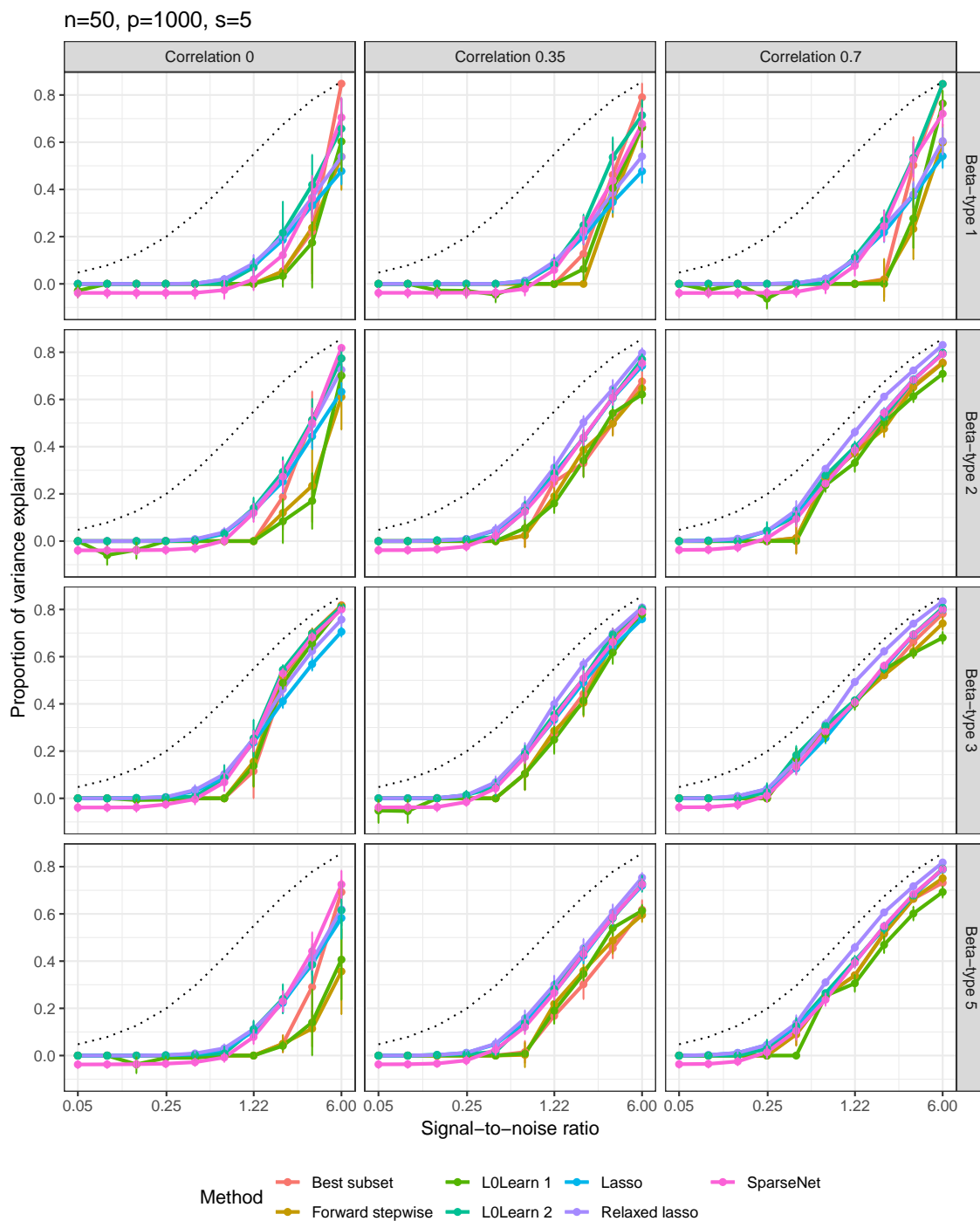
D.3.1 Relative risk (to null model)



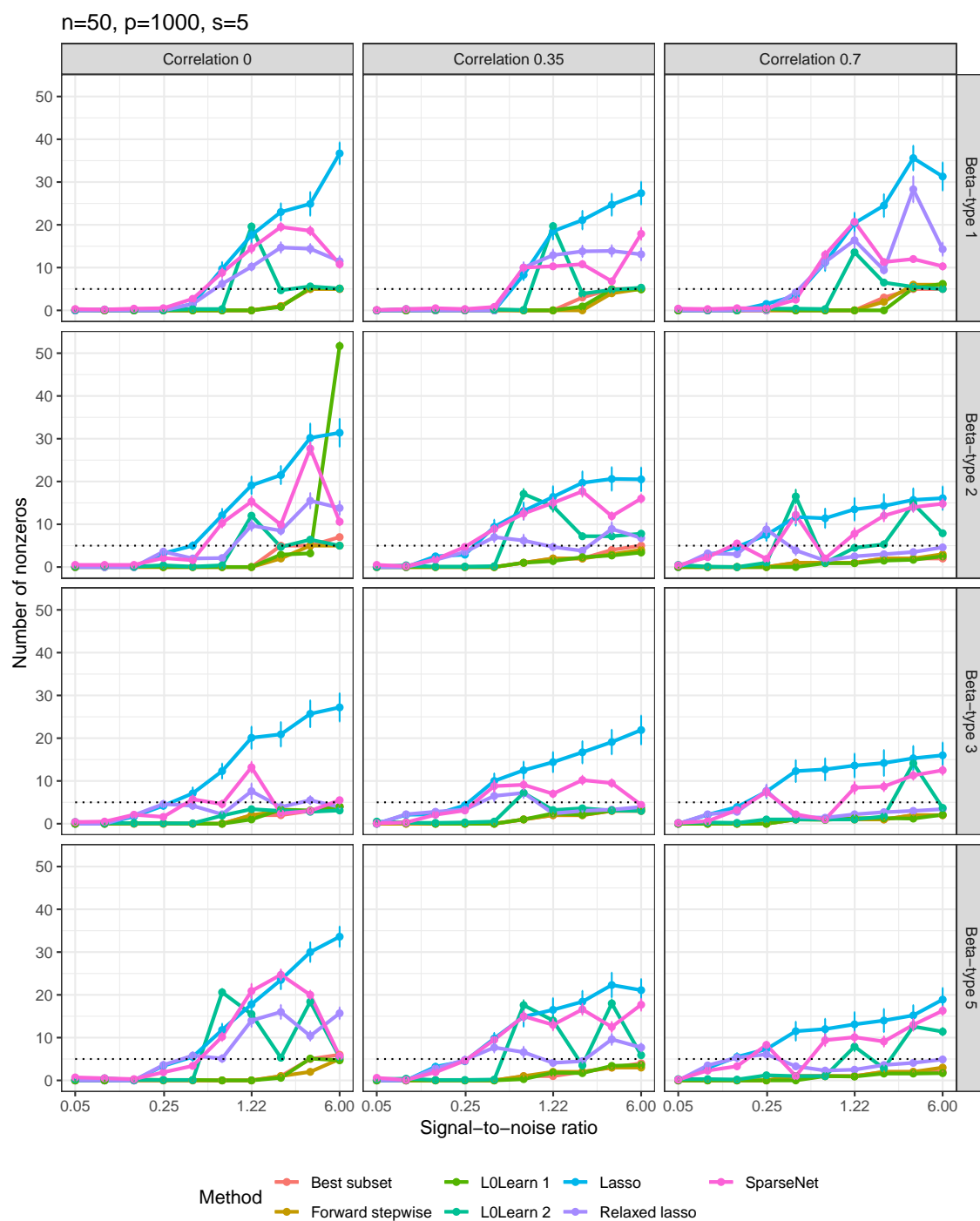
D.3.2 Relative test error (to Bayes)



D.3.3 Proportion of variance explained

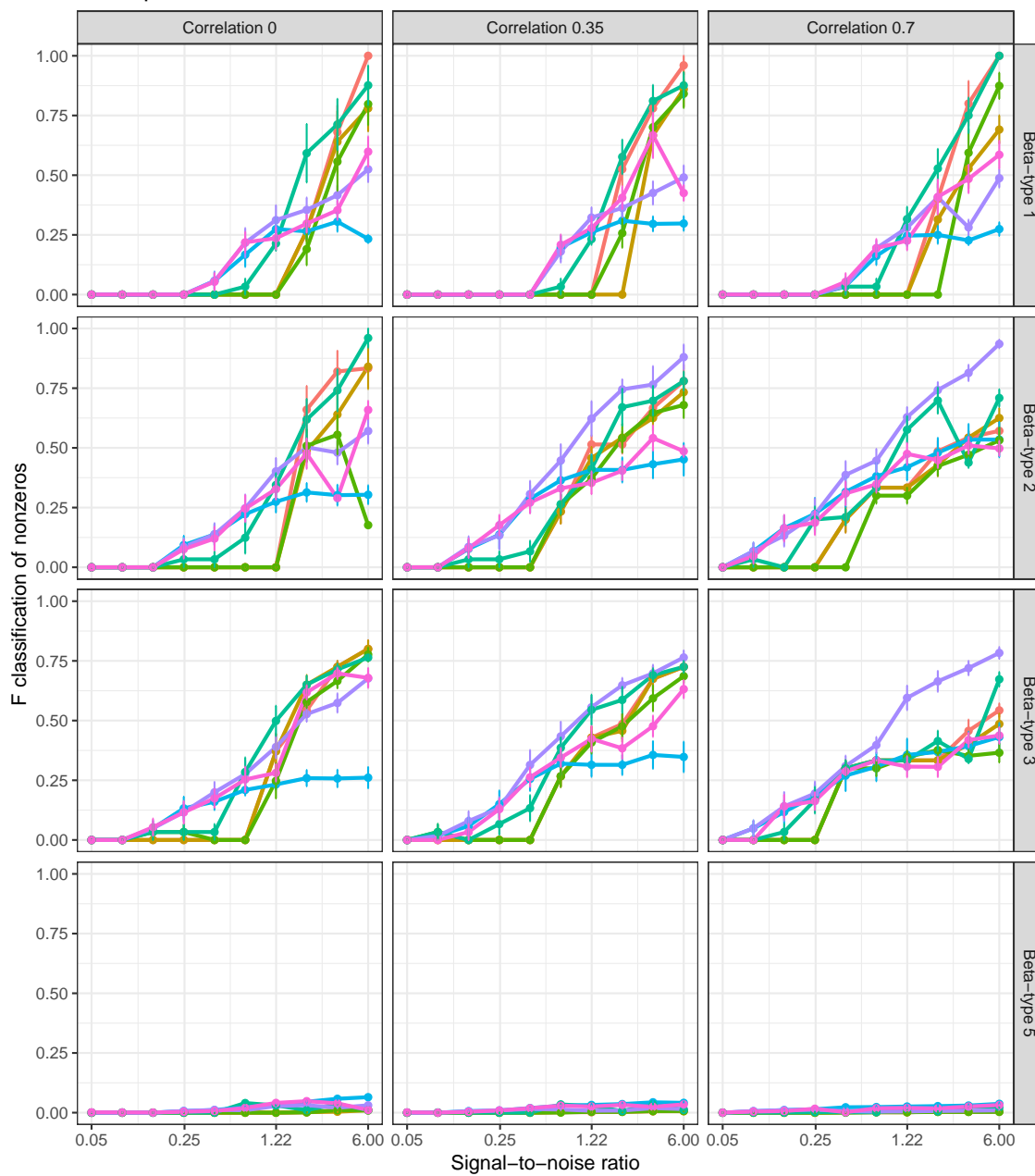


D.3.4 Number of nonzero coefficients



D.3.5 F-score

$n=50, p=1000, s=5$

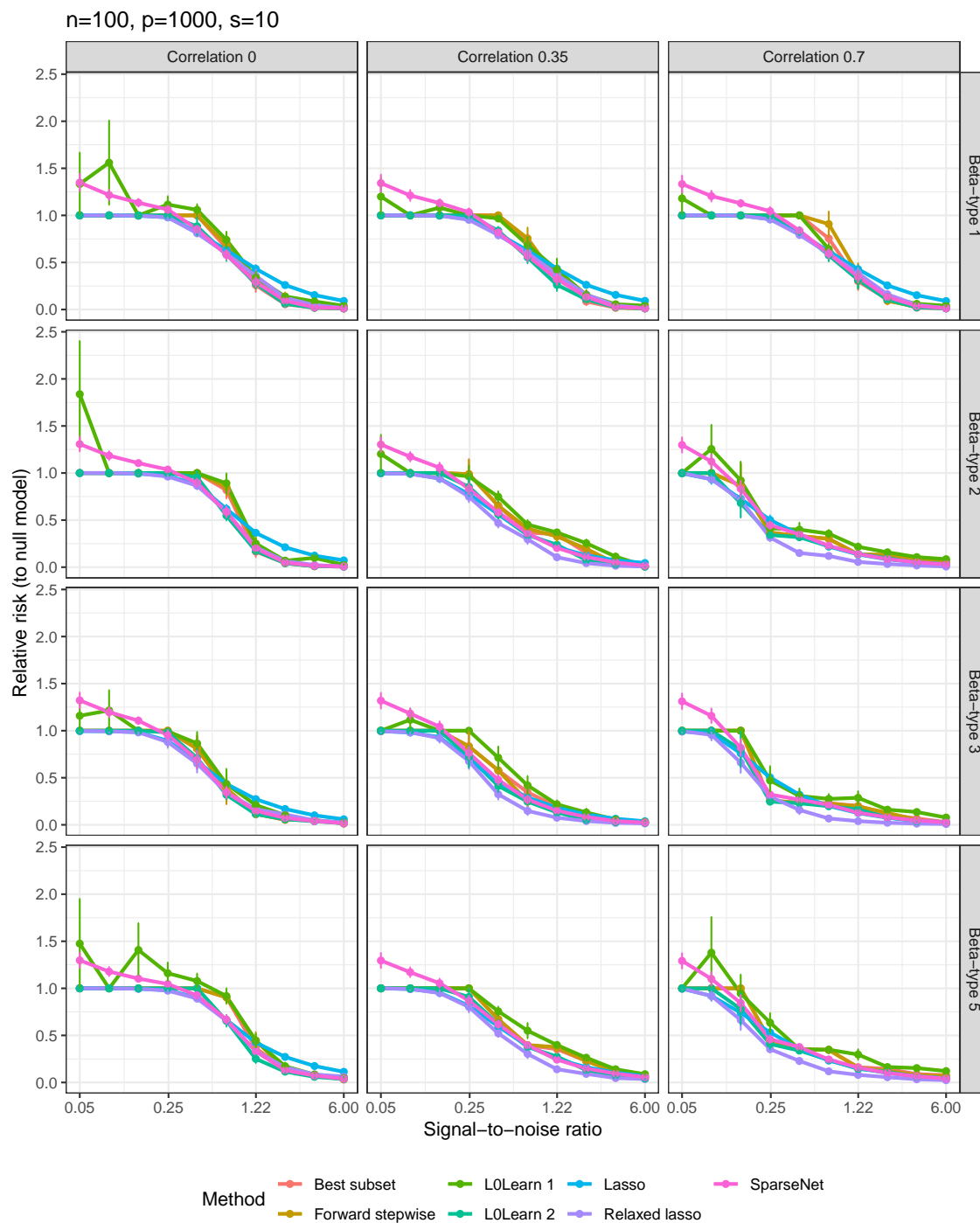


Method

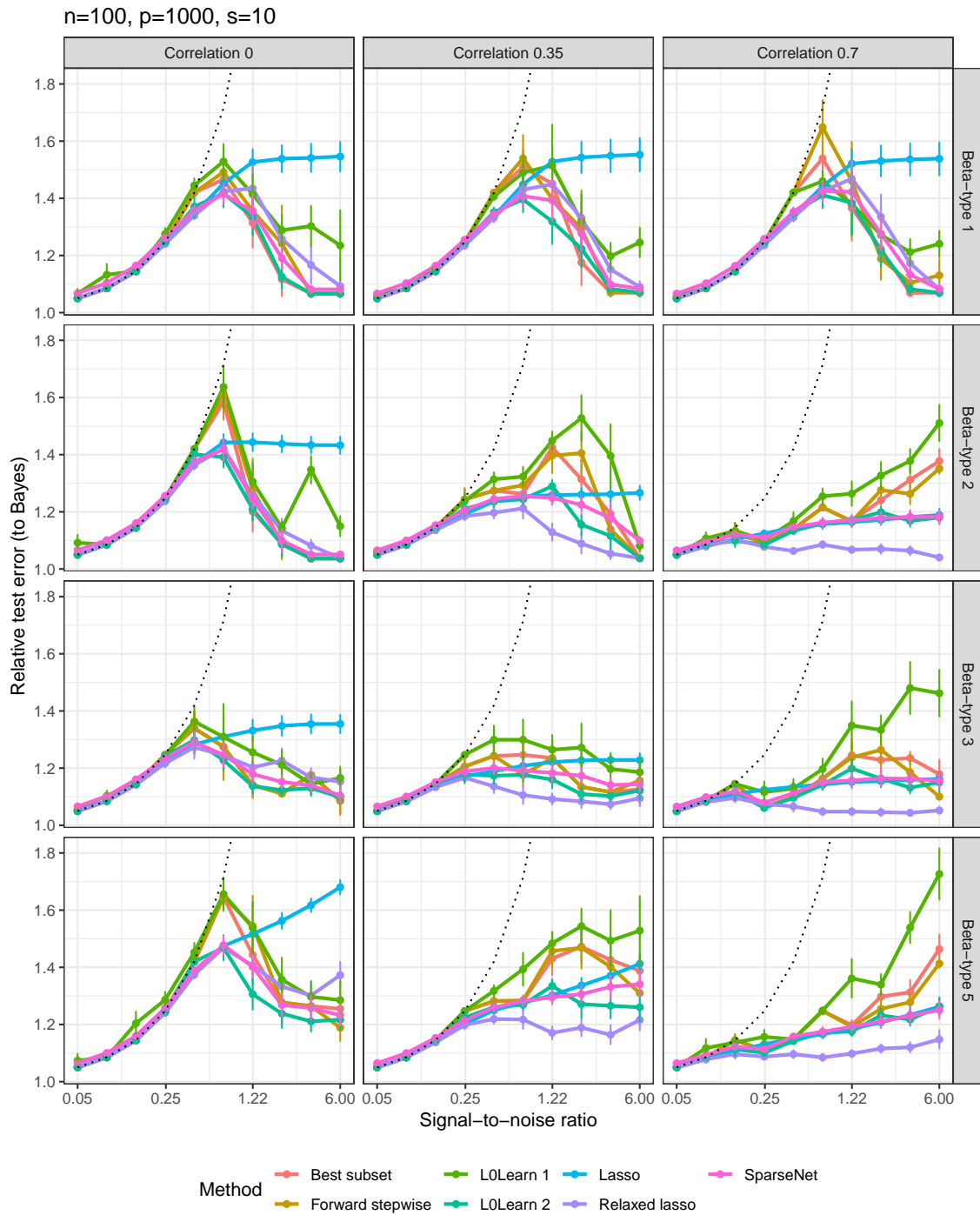
- Best subset
- L0Learn 1
- Lasso
- SparseNet
- Forward stepwise
- L0Learn 2
- Relaxed lasso

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

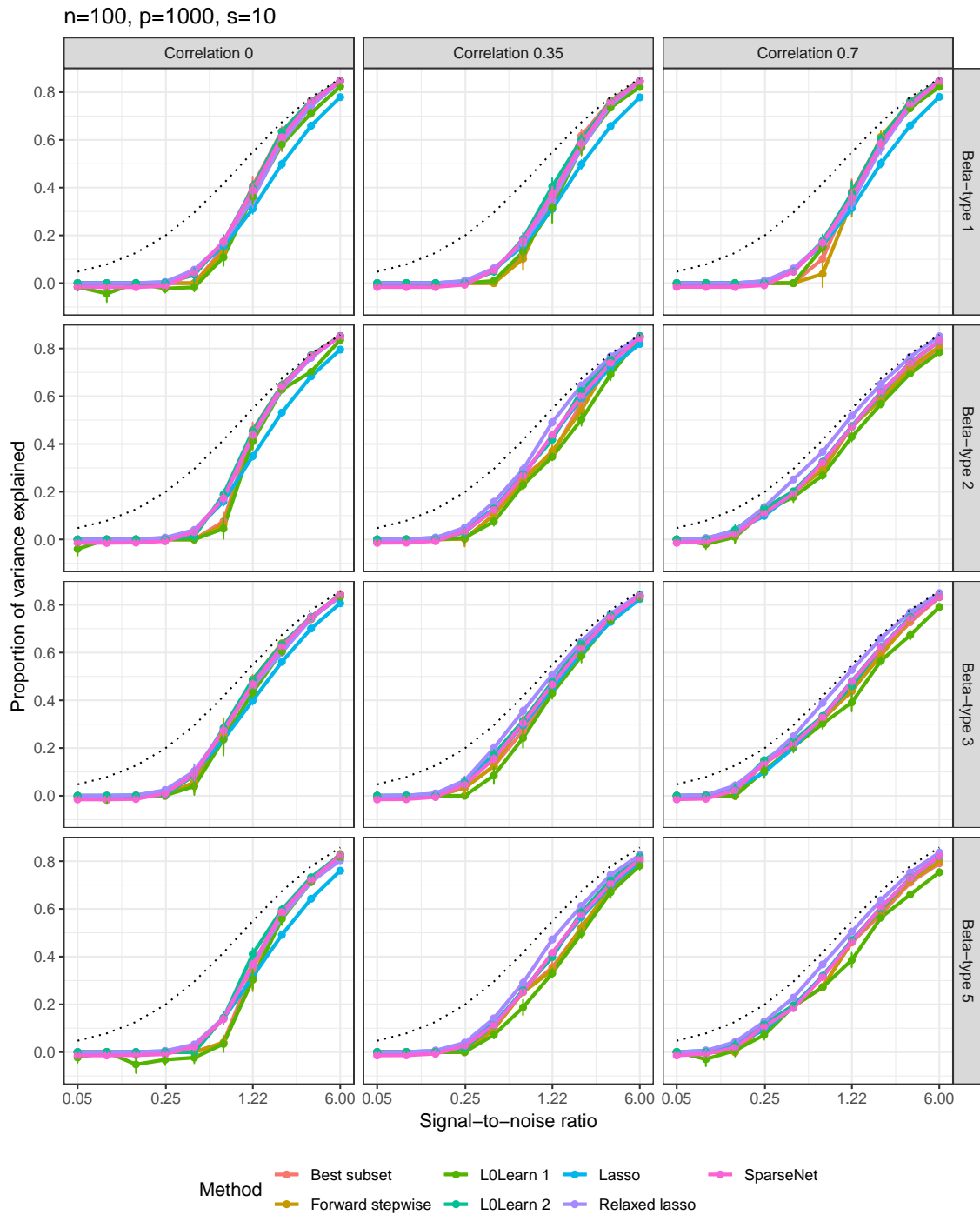
D.4.1 Relative risk (to null model)



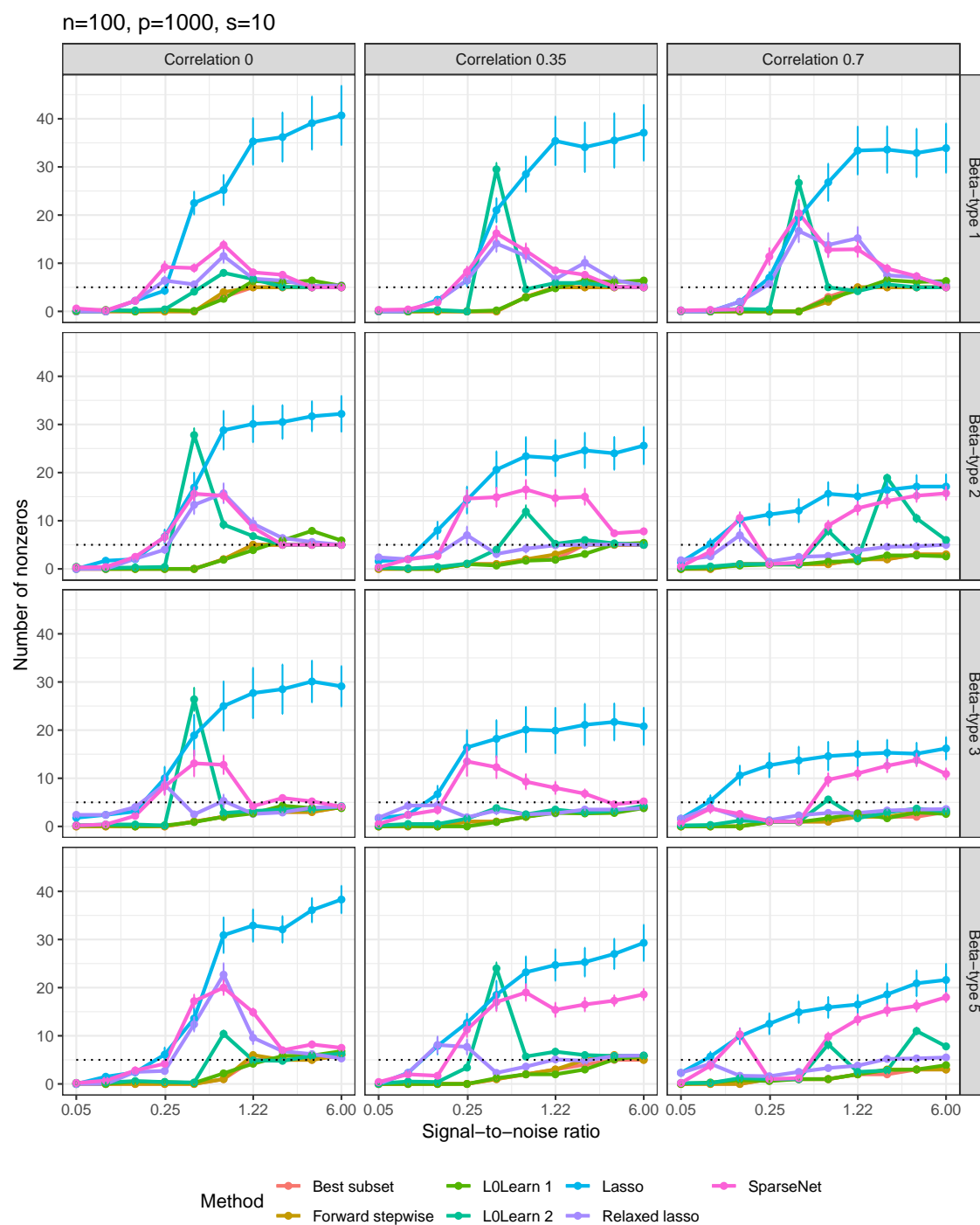
D.4.2 Relative test error (to Bayes)



D.4.3 Proportion of variance explained

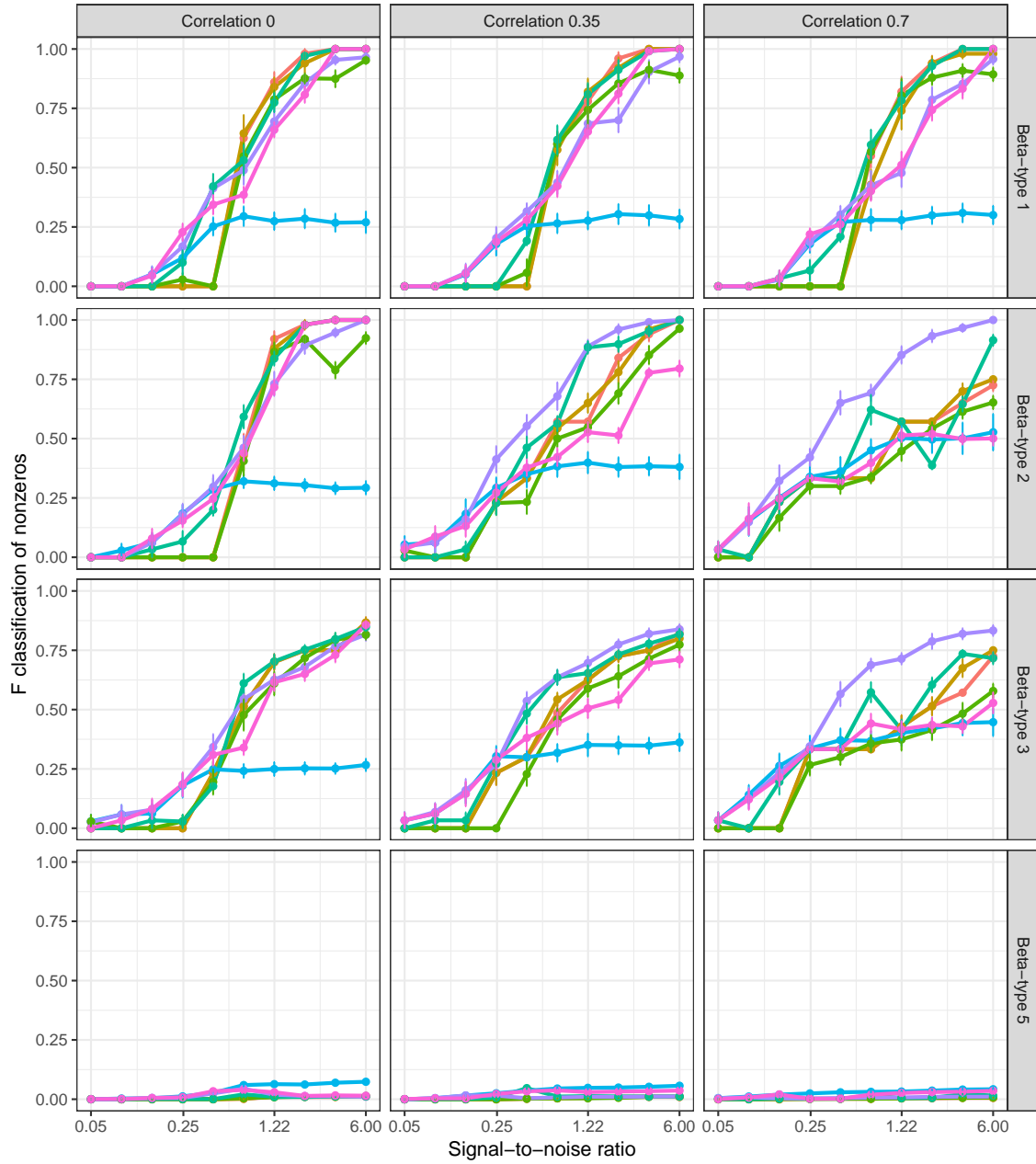


D.4.4 Number of nonzero coefficients



D.4.5 F-score

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



Method

- Best subset
- L0Learn 1
- Lasso
- SparseNet
- Forward stepwise
- L0Learn 2
- Relaxed lasso