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 $\ell_0$ penalty), L0Learn 2 (mixture of $\ell_0$ and $\ell_1$ penalties), and SparseNet.
Contents

A Validation tuning

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

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
<table>
<thead>
<tr>
<th>Subsection</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.4.3</td>
<td>Proportion of variance explained</td>
<td>82</td>
</tr>
<tr>
<td>D.4.4</td>
<td>Number of nonzero coefficients</td>
<td>83</td>
</tr>
<tr>
<td>D.4.5</td>
<td>F-score</td>
<td>84</td>
</tr>
</tbody>
</table>
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)

n=100, p=10, s=5

<table>
<thead>
<tr>
<th>Method</th>
<th>Beta-type 1</th>
<th>Beta-type 2</th>
<th>Beta-type 3</th>
<th>Beta-type 4</th>
<th>Beta-type 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward stepwise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Signal-to-noise ratio

Method: Best subset, Forward stepwise, Lasso, Relaxed lasso
A.1.3 Proportion of variance explained

\[ n=100, p=10, s=5 \]

<table>
<thead>
<tr>
<th>Correlation 0</th>
<th>Correlation 0.35</th>
<th>Correlation 0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta-type 1</td>
<td>Beta-type 2</td>
<td>Beta-type 3</td>
</tr>
<tr>
<td>Beta-type 2</td>
<td>Beta-type 4</td>
<td>Beta-type 5</td>
</tr>
</tbody>
</table>

Signal-to-noise ratio

Proportion of variance explained

<table>
<thead>
<tr>
<th>Method</th>
<th>Best subset</th>
<th>Forward stepwise</th>
<th>Lasso</th>
<th>Relaxed lasso</th>
</tr>
</thead>
</table>

Correlation 0

Correlation 0.35

Correlation 0.7

Method: Best subset, Forward stepwise, Lasso, Relaxed lasso
A.1.4 Number of nonzero coefficients

n=100, p=10, s=5

Method
- Best subset
- Forward stepwise
- Lasso
- Relaxed lasso

Signal-to-noise ratio

Number of nonzeros
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)

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

**Method**
- Best subset
- Forward stepwise
- Lasso
- Relaxed lasso

**Signal-to-noise ratio**

**Beta-type**
- Beta-type 1
- Beta-type 2
- Beta-type 3
- Beta-type 4
- Beta-type 5
A.2.3 Proportion of variance explained

![Graph showing the proportion of variance explained for different methods and signal-to-noise ratios.

Method: Best subset, Forward stepwise, Lasso, Relaxed lasso.

n=500, p=100, s=5

- Correlation 0
- Correlation 0.35
- Correlation 0.7

- Beta-type 1
- Beta-type 2
- Beta-type 3
- Beta-type 4
- Beta-type 5

Signal-to-noise ratio

Proportion of variance explained

0.00 0.25 0.50 0.75

0.05 0.25 0.50 0.75

0.05 0.25 0.50 0.75

0.05 0.25 0.50 0.75

0.05 0.25 0.50 0.75

0.05 0.25 0.50 0.75

0.05 0.25 0.50 0.75

0.05 0.25 0.50 0.75

0.05 0.25 0.50 0.75

0.05 0.25 0.50 0.75

0.05 0.25 0.50 0.75

0.05 0.25 0.50 0.75

0.05 0.25 0.50 0.75

0.05 0.25 0.50 0.75

0.05 0.25 0.50 0.75
A.2.4 Number of nonzero coefficients

\[ n=500, \ p=100, \ s=5 \]

<table>
<thead>
<tr>
<th>Method</th>
<th>Beta-type 1</th>
<th>Beta-type 2</th>
<th>Beta-type 3</th>
<th>Beta-type 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward stepwise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Method colors:
- Red: Best subset
- Green: Forward stepwise
- Blue: Lasso
- Purple: Relaxed lasso
A.2.5 F-score

n=100, p=10, s=5

Correlation 0
Correlation 0.35
Correlation 0.7

F classification of nonzeros

Method
Best subset
Forward stepwise
Lasso
Relaxed lasso

Signal-to-noise ratio
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)

![Graph showing the relative test error for different methods and signal-to-noise ratios]

**Method**
- Best subset
- Forward stepwise
- Lasso
- Relaxed lasso

**Parameters**
- n=50, p=1000, s=5
A.3.3 Proportion of variance explained

n=50, p=1000, s=5

<table>
<thead>
<tr>
<th>Method</th>
<th>Correlation 0</th>
<th>Correlation 0.35</th>
<th>Correlation 0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta-type 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Method: Best subset, Forward stepwise, Lasso, Relaxed lasso

Signal-to-noise ratio

Proportion of variance explained
A.3.4 Number of nonzero coefficients

n=50, p=1000, s=5

<table>
<thead>
<tr>
<th>Method</th>
<th>Best subset</th>
<th>Forward stepwise</th>
<th>Lasso</th>
<th>Relaxed lasso</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation 0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation 0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A.3.5 F-score

n=50, p=1000, s=5

<table>
<thead>
<tr>
<th>Correlation 0</th>
<th>Correlation 0.35</th>
<th>Correlation 0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta-type 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Method: Best subset, Forward stepwise, Lasso, Relaxed lasso

Signal-to-noise ratio vs. F classification of nonzeros
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)

\[ n=100, \ p=1000, \ s=10 \]

<table>
<thead>
<tr>
<th>Method</th>
<th>Correlation 0</th>
<th>Correlation 0.35</th>
<th>Correlation 0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td>[\text{Signal-to-noise ratio}]</td>
<td>[\text{Relative test error (to Bayes)}]</td>
<td>[\text{Beta-type 1}]</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.25</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>1.25</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>1.50</td>
<td>2.00</td>
</tr>
</tbody>
</table>

Method colors:
- **Best subset**
- **Forward stepwise**
- **Lasso**
- **Relaxed lasso**
A.4.3 Proportion of variance explained

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

Method

- Best subset
- Forward stepwise
- Lasso
- Relaxed lasso

Signal-to-noise ratio

Proportion of variance explained
A.4.4 Number of nonzero coefficients

n=100, p=1000, s=10

Method
- Best subset
- Forward stepwise
- Lasso
- Relaxed lasso

Correlation 0

Correlation 0.35

Correlation 0.7

Number of nonzeros

Signal-to-noise ratio
### A.4.5 F-score

For the given data set with $n=100$, $p=1000$, and $s=10$, the F-score was calculated for different correlation levels: 0, 0.35, and 0.7. The plots illustrate the F-classification of nonzeros for various signal-to-noise ratios, with five different Beta types. The methods considered are Best subset, Forward stepwise, Lasso, and Relaxed lasso.

#### Signal-to-Noise Ratio

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Best subset</td>
</tr>
<tr>
<td></td>
<td>Forward stepwise</td>
</tr>
<tr>
<td></td>
<td>Lasso</td>
</tr>
<tr>
<td></td>
<td>Relaxed lasso</td>
</tr>
</tbody>
</table>

#### F-Classification of Nonzeros

The graphs show the performance of each method across different signal-to-noise ratios, with Beta-type 1, Beta-type 2, Beta-type 3, and Beta-type 5.
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)

n=100, p=10, s=5

Method

Best subset
Forward stepwise
Lasso
Relaxed lasso

Relative test error (to Bayes)
B.1.3 Proportion of variance explained

\[ n=100, \ p=10, \ s=5 \]

<table>
<thead>
<tr>
<th>Method</th>
<th>Correlation 0</th>
<th>Correlation 0.35</th>
<th>Correlation 0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward stepwise</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Signal-to-noise ratio**

Method: Red = Best subset, Green = Forward stepwise, Blue = Lasso, Purple = Relaxed lasso
B.1.4 Number of nonzero coefficients

<table>
<thead>
<tr>
<th>n=100, p=10, s=5</th>
<th>Correlation 0</th>
<th>Correlation 0.35</th>
<th>Correlation 0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beta-type 1</td>
<td>Beta-type 2</td>
<td>Beta-type 3</td>
</tr>
<tr>
<td></td>
<td>Beta-type 4</td>
<td>Beta-type 5</td>
<td></td>
</tr>
</tbody>
</table>

Method: Best subset, Forward stepwise, Lasso, Relaxed lasso
B.1.5 F-score

$n=100, p=10, s=5$

<table>
<thead>
<tr>
<th>Method</th>
<th>Best subset</th>
<th>Forward stepwise</th>
<th>Lasso</th>
<th>Relaxed lasso</th>
</tr>
</thead>
</table>

Correlation 0

Correlation 0.35

Correlation 0.7

Signal-to-noise ratio

F classification of nonzeros

Beta-type 1

Beta-type 2

Beta-type 3

Beta-type 5
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)

<table>
<thead>
<tr>
<th>Method</th>
<th>Beta-type 1</th>
<th>Beta-type 2</th>
<th>Beta-type 3</th>
<th>Beta-type 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=500, p=100, s=5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation 0</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td>Correlation 0.35</td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
</tr>
<tr>
<td>Correlation 0.7</td>
<td><img src="image9" alt="Graph" /></td>
<td><img src="image10" alt="Graph" /></td>
<td><img src="image11" alt="Graph" /></td>
<td><img src="image12" alt="Graph" /></td>
</tr>
</tbody>
</table>

Method: Best subset, Forward stepwise, Lasso, Relaxed lasso

Signal-to-noise ratio

Relative test error (to Bayes)
B.2.3 Proportion of variance explained

\[
\begin{array}{cccc}
\text{Correlation 0} & \text{Correlation 0.35} & \text{Correlation 0.7} \\
0.00 & 0.00 & 0.00 \\
0.25 & 0.25 & 0.25 \\
0.50 & 0.50 & 0.50 \\
0.75 & 0.75 & 0.75 \\
1.00 & 1.00 & 1.00 \\
\end{array}
\]

\[\text{Signal-to-noise ratio} \]

\[
\begin{array}{cccc}
\text{Beta-type 1} & \text{Beta-type 2} & \text{Beta-type 3} & \text{Beta-type 5} \\
0.00 & 0.00 & 0.00 & 0.00 \\
0.25 & 0.25 & 0.25 & 0.25 \\
0.50 & 0.50 & 0.50 & 0.50 \\
0.75 & 0.75 & 0.75 & 0.75 \\
1.00 & 1.00 & 1.00 & 1.00 \\
\end{array}
\]

\[\text{n=500, p=100, s=5}\]
B.2.4 Number of nonzero coefficients

n=500, p=100, s=5

<table>
<thead>
<tr>
<th>Method</th>
<th>Beta-type 1</th>
<th>Beta-type 2</th>
<th>Beta-type 3</th>
<th>Beta-type 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward stepwise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Method: Best subset, Forward stepwise, Lasso, Relaxed lasso

Signal-to-noise ratio

Number of nonzeros

Correlation 0

Correlation 0.35

Correlation 0.7

0.05 0.25 0.5 0.7 1.22 6.00

0 10 20 30
B.2.5 F-score

For the given data set with parameters $n=500$, $p=100$, $s=5$, the F-score is plotted against the signal-to-noise ratio for different correlation levels (Correlation 0, Correlation 0.35, Correlation 0.7) and β-types (Beta-type 1, Beta-type 2, Beta-type 3, Beta-type 5).

The plots show the performance of different methods for F-classification of nonzeros:
- Best subset
- Forward stepwise
- Lasso
- Relaxed lasso

The methods are represented by different colors on the plot.
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)

\[ n=50, p=1000, s=5 \]

<table>
<thead>
<tr>
<th></th>
<th>Correlation 0</th>
<th>Correlation 0.35</th>
<th>Correlation 0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Method: Best subset, Forward stepwise, Lasso, Relaxed lasso

Signal-to-noise ratio

Relative test error (to Bayes)
B.3.3 Proportion of variance explained

\[
\begin{array}{c|c|c|c|c}
\text{n=50, p=1000, s=5} & \text{Correlation 0} & \text{Correlation 0.35} & \text{Correlation 0.7} \\
\hline
\text{Method} & \text{Best subset} & \text{Forward stepwise} & \text{Lasso} & \text{Relaxed lasso} \\
\end{array}
\]
B.3.4 Number of nonzero coefficients

n=50, p=1000, s=5

<table>
<thead>
<tr>
<th>Method</th>
<th>Best subset</th>
<th>Forward stepwise</th>
<th>Lasso</th>
<th>Relaxed lasso</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation 0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation 0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Signal-to-noise ratio
Number of nonzeros

Method: Best subset, Forward stepwise, Lasso, Relaxed lasso
### B.3.5 F-score

For $n=50$, $p=1000$, $s=5$, the graphs illustrate the performance of different methods in terms of the signal-to-noise ratio and the F-score classification of nonzeros. The methods include Best subset, Forward stepwise, Lasso, and Relaxed lasso. The graphs are categorized by Beta-type (1 to 5), each showing a range of correlation values (0, 0.35, 0.7). The x-axis represents the signal-to-noise ratio, while the y-axis shows the F-score classification of nonzeros.
B.4 **High-10 setting**: \( n = 100, \ p = 1000, \ s = 10 \)

B.4.1 **Relative risk (to null model)**

```
<table>
<thead>
<tr>
<th>Method</th>
<th>Beta-type 1</th>
<th>Beta-type 2</th>
<th>Beta-type 3</th>
<th>Beta-type 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward stepwise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Relative risk (to null model)
B.4.2 Relative test error (to Bayes)

\[ n=100, \ p=1000, \ s=10 \]

<table>
<thead>
<tr>
<th>Method</th>
<th>Best subset</th>
<th>Forward stepwise</th>
<th>Lasso</th>
<th>Relaxed lasso</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Beta-type 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Beta-type 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Beta-type 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Beta-type 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Signal-to-noise ratio

Relative test error (to Bayes)
B.4.3 Proportion of variance explained

\[ \text{Method} \quad \text{Best subset} \quad \text{Forward stepwise} \quad \text{Lasso} \quad \text{Relaxed lasso} \]

\[ \text{n}=100, \text{p}=1000, \text{s}=10 \]
B.4.4 Number of nonzero coefficients

\[ \text{Correlation 0} \]
\[ \text{Correlation 0.35} \]
\[ \text{Correlation 0.7} \]

\( n=100, p=1000, s=10 \)

<table>
<thead>
<tr>
<th>Method</th>
<th>Best subset</th>
<th>Forward stepwise</th>
<th>Lasso</th>
<th>Relaxed lasso</th>
</tr>
</thead>
</table>

Method colors:
- Red: Best subset
- Green: Forward stepwise
- Cyan: Lasso
- Purple: Relaxed lasso

Signal-to-noise ratio
Number of nonzeros

\[ \text{Signal-to-noise ratio} \]
\[ \text{Number of nonzeros} \]

\( 0 \to 40 \)
\( 0 \to 40 \)
\( 0 \to 40 \)

\( 0.05 \to 6.00 \)
\( 0.05 \to 6.00 \)
\( 0.05 \to 6.00 \)
B.4.5 F-score

n=100, p=1000, s=10

Correlation 0

Correlation 0.35

Correlation 0.7

Beta-type 1

Beta-type 2

Beta-type 3

Beta-type 5

Signal-to-noise ratio

F classification of nonzeros

Method

Best subset
Forward stepwise
Lasso
Relaxed lasso

n=100, p=1000, s=10

Method

Best subset
Forward stepwise
Lasso
Relaxed lasso
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)

n=100, p=10, s=5

<table>
<thead>
<tr>
<th>Method</th>
<th>Signal-to-noise ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td></td>
</tr>
<tr>
<td>L0Learn 1</td>
<td></td>
</tr>
<tr>
<td>L0Learn 2</td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td></td>
</tr>
<tr>
<td>Forward stepwise</td>
<td></td>
</tr>
<tr>
<td>SparseNet</td>
<td></td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td></td>
</tr>
</tbody>
</table>

Method legend:
- Best subset
- L0Learn 1
- L0Learn 2
- Lasso
- Forward stepwise
- SparseNet
- Relaxed lasso
C.1.3 Proportion of variance explained

n=100, p=10, s=5
C.1.4 Number of nonzero coefficients

\[ n=100, p=10, s=5 \]

\[ \text{Correlation 0} \]

\[ \text{Correlation 0.35} \]

\[ \text{Correlation 0.7} \]

\[ \text{Number of nonzeros} \]

\[ \text{Signal-to-noise ratio} \]

\[ \text{Method} \]

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

48
C.1.5 F-score

![Graph showing F-score comparison for various methods and signal-to-noise ratios. The legend indicates the methods: Best subset, L0Learn 1, L0Learn 2, Lasso, SparseNet, Forward stepwise, Relaxed lasso. The graphs are color-coded to differentiate between methods.](image-url)
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)

n=500, p=100, s=5

<table>
<thead>
<tr>
<th>Method</th>
<th>Method</th>
<th>Method</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta−type 1</td>
<td>Beta−type 2</td>
<td>Beta−type 3</td>
<td>Beta−type 5</td>
</tr>
<tr>
<td>Best subset</td>
<td>L0Learn 1</td>
<td>Lasso</td>
<td>SparseNet</td>
</tr>
<tr>
<td>Forward stepwise</td>
<td>L0Learn 2</td>
<td>Relaxed lasso</td>
<td></td>
</tr>
</tbody>
</table>

Signal–to–noise ratio

Correlation 0

Correlation 0.35

Correlation 0.7
C.2.3 Proportion of variance explained

\[ n=500, \, p=100, \, s=5 \]

<table>
<thead>
<tr>
<th>Correlation 0</th>
<th>Correlation 0.35</th>
<th>Correlation 0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta-type 1</td>
<td>Beta-type 2</td>
<td>Beta-type 3</td>
</tr>
<tr>
<td>Beta-type 4</td>
<td>Beta-type 5</td>
<td></td>
</tr>
</tbody>
</table>

Signal-to-noise ratio

Method
- Best subset
- L0Learn 1
- L0Learn 2
- Lasso
- Forward stepwise
- SparseNet
- Relaxed lasso
C.2.4 Number of nonzero coefficients

n=500, p=100, s=5

- Correlation 0
- Correlation 0.35
- Correlation 0.7

Number of nonzeros vs. Signal-to-noise ratio for different methods:
- Best subset
- L0Learn 1
- L0Learn 2
- Lasso
- SparseNet
- Forward stepwise
- Relaxed lasso

Beta-type 1
Beta-type 2
Beta-type 3
Beta-type 5
C.2.5 F-score

n=100, p=10, s=5

<table>
<thead>
<tr>
<th>Method</th>
<th>Beta-type 1</th>
<th>Beta-type 2</th>
<th>Beta-type 3</th>
<th>Beta-type 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0Learn 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SparseNet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward stepwise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0Learn 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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)

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Beta−type 1</th>
<th>Beta−type 2</th>
<th>Beta−type 3</th>
<th>Beta−type 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0Learn 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0Learn 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SparseNet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward stepwise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
C.3.3 Proportion of variance explained

n=50, p=1000, s=5

<table>
<thead>
<tr>
<th>Method</th>
<th>Signal-to-noise ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td>0.05, 0.25, 1.22, 6.00</td>
</tr>
<tr>
<td>L0Learn 1</td>
<td>0.05, 0.25, 1.22, 6.00</td>
</tr>
<tr>
<td>L0Learn 2</td>
<td>0.05, 0.25, 1.22, 6.00</td>
</tr>
<tr>
<td>L0Learn 3</td>
<td>0.05, 0.25, 1.22, 6.00</td>
</tr>
<tr>
<td>L0Learn 4</td>
<td>0.05, 0.25, 1.22, 6.00</td>
</tr>
<tr>
<td>L0Learn 5</td>
<td>0.05, 0.25, 1.22, 6.00</td>
</tr>
<tr>
<td>Forward stepwise</td>
<td>0.05, 0.25, 1.22, 6.00</td>
</tr>
<tr>
<td>Lasso</td>
<td>0.05, 0.25, 1.22, 6.00</td>
</tr>
<tr>
<td>SparseNet</td>
<td>0.05, 0.25, 1.22, 6.00</td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td>0.05, 0.25, 1.22, 6.00</td>
</tr>
</tbody>
</table>
C.3.4 Number of nonzero coefficients

\[ n=50, p=1000, s=5 \]

<table>
<thead>
<tr>
<th>Method</th>
<th>Beta-type 1</th>
<th>Beta-type 2</th>
<th>Beta-type 3</th>
<th>Beta-type 4</th>
<th>Beta-type 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0Learn 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0Learn 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SparseNet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward stepwise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
C.3.5 F-score

n=50, p=1000, s=5

<table>
<thead>
<tr>
<th>Method</th>
<th>Beta-type 1</th>
<th>Beta-type 2</th>
<th>Beta-type 3</th>
<th>Beta-type 4</th>
<th>Beta-type 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0Learn 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0Learn 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0Learn 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0Learn 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0Learn 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward stepwise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SparseNet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
C.4 High-10 setting: $n = 100$, $p = 1000$, $s = 10$

C.4.1 Relative risk (to null model)

![Graph showing relative risk to null model for different correlation levels and signal-to-noise ratios.](image)

- **Method**:
  - Best subset
  - L0Learn 1
  - L0Learn 2
  - Lasso
  - Relaxed lasso
  - Forward stepwise

- **Correlation Levels**:
  - 0
  - 0.35
  - 0.7

- **Signal-to-noise ratio**:
  - $n = 100$, $p = 1000$, $s = 10$
C.4.2 Relative test error (to Bayes)

\[ n=100, \ p=1000, \ s=10 \]

<table>
<thead>
<tr>
<th>Method</th>
<th>Signal-to-noise ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td>0.05 0.25 1.22 6.00</td>
</tr>
<tr>
<td>L0Learn 1</td>
<td>0.05 0.25 1.22 6.00</td>
</tr>
<tr>
<td>Lasso</td>
<td>0.05 0.25 1.22 6.00</td>
</tr>
<tr>
<td>SparseNet</td>
<td>0.05 0.25 1.22 6.00</td>
</tr>
<tr>
<td>Forward stepwise</td>
<td>0.05 0.25 1.22 6.00</td>
</tr>
<tr>
<td>L0Learn 2</td>
<td>0.05 0.25 1.22 6.00</td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td>0.05 0.25 1.22 6.00</td>
</tr>
</tbody>
</table>

Method: Best subset, L0Learn 1, Lasso, SparseNet, Forward stepwise, L0Learn 2, Relaxed lasso
### C.4.3 Proportion of variance explained

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Signal-to-noise ratio</th>
<th>Proportion of variance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td>0.05</td>
<td>0.0</td>
</tr>
<tr>
<td>L0Learn 1</td>
<td>0.25</td>
<td>0.2</td>
</tr>
<tr>
<td>Lasso</td>
<td>0.25</td>
<td>0.4</td>
</tr>
<tr>
<td>SparseNet</td>
<td>0.25</td>
<td>0.2</td>
</tr>
<tr>
<td>Forward stepwise</td>
<td>0.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

![Graph showing Proportion of variance explained for different signal-to-noise ratios and methods](image-url)
C.4.4 Number of nonzero coefficients

![Graph showing number of nonzero coefficients for different methods and signal-to-noise ratios](image-url)

- **n=100, p=1000, s=10**

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

- **Beta-type 1**
- **Beta-type 2**
- **Beta-type 3**
- **Beta-type 4**
- **Beta-type 5**
C.4.5 F-score

n=100, p=1000, s=10

<table>
<thead>
<tr>
<th>Method</th>
<th>Best subset</th>
<th>L0Learn 1</th>
<th>Lasso</th>
<th>SparseNet</th>
<th>Forward stepwise</th>
<th>L0Learn 2</th>
<th>Relaxed lasso</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta-type 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Signal-to-noise ratio
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)

n=100, p=10, s=5
D.1.3 Proportion of variance explained

n=100, p=10, s=5

<table>
<thead>
<tr>
<th>Method</th>
<th>Beta-type 1</th>
<th>Beta-type 2</th>
<th>Beta-type 3</th>
<th>Beta-type 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0Learn 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward stepwise</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L0Learn 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SparseNet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Method colors:
- Best subset
- L0Learn 1
- L0Learn 2
- Lasso
- Relaxed lasso
- Forward stepwise
- SparseNet
D.1.4 Number of nonzero coefficients

n=100, p=10, s=5

Number of nonzeros

Signal-to-noise ratio

Method
- Best subset
- L0Learn 1
- L0Learn 2
- Lasso
- SparseNet
- Forward stepwise
- Relaxed lasso
D.1.5 F-score

\[ n=100, p=10, s=5 \]

[Graph showing F-score for different methods and signal-to-noise ratios for different correlation levels (0, 0.35, 0.7). The methods include Best subset, L0Learn 1, L0Learn 2, Lasso, Relaxed lasso, Forward stepwise, and SparseNet. Each method is represented by a different line color and marker.]
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)

n=500, p=100, s=5

Signal-to-noise ratio

Method
- Best subset
- L0Learn 1
- L0Learn 2
- Lasso
- SparseNet
- Forward stepwise
- Relaxed lasso
D.2.3 Proportion of variance explained

\[ n=500, p=100, s=5 \]

<table>
<thead>
<tr>
<th>Correlation 0</th>
<th>Correlation 0.35</th>
<th>Correlation 0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta-type 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta-type 5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Method**
- Best subset
- L0Learn 1
- L0Learn 2
- SparseNet
- Forward stepwise
- Relaxed lasso

**Signal-to-noise ratio**

- 0.05
- 0.25
- 1.22
- 6.00

**Proportion of variance explained**

- 0.00
- 0.25
- 0.50
- 0.75
D.2.4 Number of nonzero coefficients

n=500, p=100, s=5
D.2.5  F-score

n=500, p=100, s=5

F classification of nonzeros

Signal-to-noise ratio

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

\[ n=50, p=1000, s=5 \]

<table>
<thead>
<tr>
<th>Method</th>
<th>D.3.2 Relative test error (to Bayes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best subset</td>
<td></td>
</tr>
<tr>
<td>L0Learn 1</td>
<td></td>
</tr>
<tr>
<td>Lasso</td>
<td></td>
</tr>
<tr>
<td>SparseNet</td>
<td></td>
</tr>
<tr>
<td>Forward stepwise</td>
<td></td>
</tr>
<tr>
<td>L0Learn 2</td>
<td></td>
</tr>
<tr>
<td>Relaxed lasso</td>
<td></td>
</tr>
</tbody>
</table>

Signal-to-noise ratio

<table>
<thead>
<tr>
<th>Correlation 0</th>
<th>Correlation 0.35</th>
<th>Correlation 0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beta-type 1</td>
<td>Beta-type 2</td>
</tr>
<tr>
<td></td>
<td>Beta-type 3</td>
<td>Beta-type 5</td>
</tr>
</tbody>
</table>

Signal-to-noise ratio

Graph showing the relative test error to Bayes for different correlation levels and signal-to-noise ratios for various methods.
D.3.3 Proportion of variance explained

n=50, p=1000, s=5
D.3.4 Number of nonzero coefficients

n=50, p=1000, s=5

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

Signal-to-noise ratio

Beta-type 1

Beta-type 2

Beta-type 3

Beta-type 4

Beta-type 5
D.3.5 F-score

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

- Correlation 0
- Correlation 0.35
- Correlation 0.7

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

Signal-to-noise ratio vs. Classification of nonzero parameters
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)

n=100, p=1000, s=10

<table>
<thead>
<tr>
<th>Correlation 0</th>
<th>Correlation 0.35</th>
<th>Correlation 0.7</th>
</tr>
</thead>
</table>

Beta-type 1

Beta-type 2

Beta-type 3

Beta-type 5

Method:
- Best subset
- L0Learn 1
- L0Learn 2
- Lasso
- Relaxed lasso
- Forward stepwise

Signal-to-noise ratio
D.4.3 Proportion of variance explained

\begin{tabular}{|c|c|c|}
\hline
\textbf{n=100, p=1000, s=10} & \textbf{Correlation 0} & \textbf{Correlation 0.35} & \textbf{Correlation 0.7} \\
\hline
\end{tabular}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{proportion_variance_explained.png}
\caption{Proportion of variance explained for different correlation levels and signal-to-noise ratios.}
\end{figure}

- \textbf{Beta-type 1}
- \textbf{Beta-type 2}
- \textbf{Beta-type 3}
- \textbf{Beta-type 5}

\textbf{Method}
\begin{itemize}
\item Best subset
\item L0Learn 1
\item L0Learn 2
\item Lasso
\item Relaxed lasso
\item Forward stepwise
\end{itemize}
D.4.4 Number of nonzero coefficients

n=100, p=1000, s=10

Correlation 0  Correlation 0.35  Correlation 0.7

Beta-type 1

Beta-type 2

Beta-type 3

Beta-type 5

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

Signal-to-noise ratio

Number of nonzero

0 0.05 0.25 1.22 6.00 0 0.05 0.25 1.22 6.00 0 0.05 0.25 1.22 6.00
D.4.5 F-score

![Graph showing F-score for different methods and signal-to-noise ratios.]

- **n=100, p=1000, s=10**

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

- **Correlation Levels**:
  - Correlation 0
  - Correlation 0.35
  - Correlation 0.7

- **Signal-to-noise Ratio**:
  - Values range from 0.05 to 6.00