Unidimensional factor models imply weaker partial correlations than zero-order correlations: Supplementary Materials

Riet van Bork, Raoul P. P. P. Grasman, Lourens Waldorp

University of Amsterdam

Unidimensional factor models imply weaker partial correlations than zero-order correlations: Supplementary Materials

Bootstrap test

The following provides R-code for the bootstrap test.

```
onefactor.test <- function(X, M = 1000, alpha=0.05) {
        cor2pcor = function(x) {x = -solve(x); diag(x) =
           abs(diag(x)); cov2cor(x)}
        s_ij = replicate(M, {
                R = cor(X[sample(nrow(X),replace=TRUE),]);
                pR = cor2pcor(R);
                abs(2 * pR[upper.tri(R)] / R[upper.tri(R)] - 1)
        })
        R = cor(X)
        pR = cor2pcor(R)
        s_ij = cbind(abs(2 * pR[upper.tri(R)] / R[upper.tri(R)]
           - 1), s_ij)
        tests = apply(s_ij, 1, function(x) {structure(list(
                statistic = structure(x[1], names = "S[i, j]"),
                parameter = structure(M, names = "number of
                   bootstrap samples"),
                conf.int = structure(quantile(x[-1],
                   c(alpha/2,1-alpha/2)), conf.level=1-alpha),
                null.value = structure(1, names="S[i, j]"),
                alternative = "greater",
                method = "Bootstrap test of S[i, j] = |2 *
                   pcor[i,j]/cor[i,j] - 1| < 1",
                data.name = deparse(substitute(X)),
                samples = x[-1]
```

```
        ), class="htest")})
        names(tests) = paste("S[", outer(1:ncol(R),1:ncol(R),
          paste, sep=", ")[upper.tri(R)],"]", sep="")
        tests
}
```

The input to the function (onefactor.test) is a data matrix (X) in which each row corresponds to one observation unit (participant, animal, etc.), an integer (M) that specifies the number of bootstrap samples (defaults to 1000), and a number between 0 and 1 (alpha) which specifies the desired significance level of the test. The output is a list of hypothesis test ('htest') objects that print the outcome of the tests to the screen.

Example I: UFM

As specified in the paper, the model implied covariance matrix of a unidimensional factor model (UFM) is of the form:

$$\boldsymbol{\Sigma} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{\Theta}. \tag{1}$$

The following provides some additional information on the simulated example in the paper in which data was generated from a UFM with factor loadings close to zero. Factor loadings were sampled from a uniform distribution over $[0.05, 0.2]$, corresponding to 6 observed variables. Each factor loading had a probability of .5 to be multiplied with -1 resulting in both positive and negative factor loadings. Θ is a diagonal matrix with $\theta_{ii} = 1 - \lambda_i^2$, such that the diagonal of $\boldsymbol{\Sigma}$ is one, resulting in a correlation matrix. We simulated 60 observations. The following includes the R-code we used to create the UFM and sample data with, as well as code to create figures of the results of the test.

```
# Example: One factor model


# Sigma according to the one factor model (0 < |lambda| <1):
lambda = sample(c(-1,1),6,replace = T)*runif(6,0.05,0.2)
Sigma = outer(lambda,lambda) + diag(1-lambda^2)
```

```
X = matrix(rnorm(length(Sigma)*10),,ncol(Sigma)) %*% chol(Sigma)
(res <- onefactor.test(X))

# Plot results:
lapply(seq_along(res), function(obj, n, i)
    {hist(log(c(1,obj[[i]]$samples)),30, axes = F, xlab = "log
    S", main = n[[i]]); axis(1); abline(v=0,col=2);
    paste(n[[i]])},obj = res,n = names(res))
```

The sample correlation matrix of the simulated dataset is presented in the upper triangle of the matrix in Table 1. The corresponding sample partial correlation matrix of the data is presented in the lower triangle of this same matrix. The results from the bootstrap test on this dataset can be found in Table 2 and Figure 1. All of the CIs include one and the UFM should thus not be rejected.

Example II: random correlation matrix

Here we provide a similar example as Example I but with data that are not generated from a unidimensional factor model but from a random correlation matrix with six variables. We created the correlation matrix by taking the cross product of two 6×6 matrices that consisted of random values drawn from a uniform distribution over [-1, 1]. We simulated 60 observations from this random correlation matrix. The following includes the R code we used to create these matrices and sample data from it.

```
# Example: Random correlation matrix

# Sigma that does (necessarily) not conform to a 1-factor model:
Sigma = cov2cor(crossprod(matrix(runif(6^2,-1,1),6)))

X = matrix(rnorm(length(Sigma)*10),,ncol(Sigma)) %*% chol(Sigma)
(res <- onefactor.test(X))
```

```
# Plot results:
lapply(seq_along(res), function(obj, n, i)
  {hist(log(c(1,obj[[i]]$samples)),30, axes = F, xlab = "log
  S", main = n[[i]]); axis(1); abline(v=0,col=2);
  paste(n[[i]])},obj = res,n = names(res))
```

The sample correlation matrix is presented in the upper triangle of the matrix in Table 3. The corresponding partial correlation matrix is presented in the lower triangle of this same matrix. The results of the bootstrap test to these data are presented in Table 4 and Figure 2. There are eight pairs of variables for which the CI does not include the value one. The sample correlation matrix and sample partial correlation matrix show that four of these pairs correspond to zero-order correlations that have a different sign than the partial correlation (r_{14}, r_{35}, r_{16} and r_{56}). The other four pairs correspond to partial correlations that are stronger than the zero-order correlation (r_{34}, r_{45}, r_{36} and r_{46}).

Simulation study: Performance of the empirical bootstrap test

We performed a simulation study to obtain the false positive rate and the power of the bootstrap test in rejecting the UFM when the underlying model is not a UFM. We rejected the unidimensional factor model when the bootstrap test identified at least one partial correlations that was not between zero and the zero-order correlation. We tested the power of the test in three different situations. That is, we tested the power of the bootstrap test (1) when the population correlation matrix is generated from a UFM but misfit is created by adding or subtracting a given value from one of the correlations, (2) when the population correlation matrix is generated from a UFM but misfit is created by switching the sign of one of the correlations in the population correlation matrix, and (3) when the population correlation matrix is a random correlation matrix.

For the alternative model in which a given value was added to a correlation, we considered three different values of delta (0.2, 0.4 and 0.6). This results in six models:

the UFM and five alternative models. For these six models we considered 3 different numbers of observed variables (5, 10 and 15 variables), 10 different sample sizes (100 to 1000 in steps of 100) and considered both positive manifold data (since it is common in psychology to have solely positive correlations), and data with both positive and negative correlations, resulting in a total of 360 conditions. We used 1000 iterations for each condition.

For conditions in which the UFM is the true data generating model, random factor loadings were drawn from a uniform distribution over [0.1, 1] (or [-1, 1] in conditions with both positive and negative correlations). We specified Θ to be a diagonal matrix with $\theta_{ii} = 1 - \lambda_i^2$, such that the diagonal of $\mathbf{\Sigma}$ is one, resulting in a correlation matrix. For the first type of alternative models a value delta (0.2, 0.4 or 0.6) was added to one randomly selected correlation in $\mathbf{\Sigma}$. In roughly half of the cases delta was multiplied by -1 so that delta was either added or subtracted. For the second type of alternative models the sign of a randomly selected correlation in $\mathbf{\Sigma}$ was switched around. For the third type of alternative models we generated random correlation matrices that were inconsistent with a UFM by standardizing the cross-product of a matrix with values randomly drawn from a uniform distribution over [-0.25, 1] (or over [-1, 1] in the conditions with both positive and negative correlations). The reason for drawing values from a uniform distribution over [-0.25, 1] rather than over [0, 1] in conditions with positive correlations, is that in the latter case the resulting correlations are very large (most correlations $> .8$). When drawing values from [-0.25, 1] the correlations have a wider range, including low correlations and sometimes even a negative correlation. We selected the matrices with only positive correlations. The R-code for generating these matrices is included below.

Results for conditions with only positive correlations

Results of the simulation study for data with only positive correlations are summarized in Table 5 and visualized in Figure 3. The false positive rate is lower than 5% for all conditions except for the condition of 15 variables and 400 observations for which the false positive rate is 5.2%. One might note that the false positive rate increases slightly with an increasing number of variables. The reason for this is that by increasing the number of variables the number of comparisons grows very fast. The Bonferroni correction corrects for this, but for a smaller α more bootstraps are needed to get the empirical distribution range over the $1 - \alpha$ interval. When the number of bootstraps is increased from 1000 to 10000, the false positive rate for 15 variables ranges between 0.8 % and 2% over the different sample sizes. It is therefore important that the number of bootstrap samples is large enough, and a smaller α requires more bootstraps.

For random correlation matrices the power of the test was always 1 except for the condition with 5 variables for which the power ranges from 0.978 to 0.999; that is, with 10 and 15 variables the test always rejected the UFM when the data was simulated from a random correlation matrix. For the other alternative models the power increased as the sample size increased.

Results for conditions with both positive and negative correlations

The results of the simulation study for data with both positive and negative correlations is summarized in Table 6 and visualized in Figure 4. The results can be interpreted in the same way as for the conditions with only positive correlations.

Specifying alternative models

The following provides R-code for generating the models that are used in the simulation study:

```r
# Alternative models in simulation study:
nV <- 5 #number of variables
lambda = runif(nV,0.1,1)
Sigma_ufm = outer(lambda,lambda) + diag(1-lambda^2) #Sigma of
    UFM


#Alternative 1: delta = 0.2 or delta = 0.4 or delta = 0.6
```

```
d <- -1
while(d<0){
        delta<-sample(c(-1,1),1)*delta
        i <- sample(1:nV,1)
        j <- sample((1:nV)[-i],1) #selecting one random
            off-diagonal element: Sigma[i,j]
        Sigma_alternative <- Sigma_ufm
        Sigma_alternative[i,j] <- Sigma_ufm[i,j]+delta
        Sigma_alternative[j,i] <- Sigma_ufm[j,i]+delta
        d <- det(Sigma_alternative)
        }

#Alternative 2: switch sign
d <- -1
while(d<0){
        i <- sample(1:nV,1)
        j <- sample((1:nV)[-i],1) #selecting one random
            off-diagonal element: Sigma[i,j]
        Sigma_alternative <- Sigma_ufm
        Sigma_alternative[i,j] <- -1*Sigma_ufm[i,j]
        Sigma_alternative[j,i] <- -1*Sigma_ufm[j,i]
        d <- det(Sigma_alternative)
        }

# Alternative 3: Random correlation matrix
#positive and negative correlations:
Sigma = cov2cor(crossprod(matrix(runif(nV^2,-1,1),nV)))
#positive correlations:
d <- 1
```

```
while(d != nV*nV){
    Sigma = cov2cor(crossprod(matrix(runif(nV^2,-0.25,1),nV)))
    d <- sum(Sigma > 0)
    }
```

	V1	V2	V3	V4	V5	V6
V1	1	0.212	0.139	0.231	0.193	0.079
V2	0.129	1	0.296	0.170	0.010	0.042
V3	0.128	0.319	1	-0.157	0.065	-0.060
V4	0.207	0.210	-0.266	1	0.193	-0.101
V5	0.158	-0.074	0.084	0.166	1	-0.088
V6	0.122	0.071	-0.110	-0.137	-0.076	1

Table 1: Upper triangle of matrix represents sample correlation matrix of dataset with 60 observations that is simulated from UFM with both positive and negative factor loadings. The absolute factor loadings come from a uniform distribution over [0.05, 0.2]. The lower triangle of the matrix represents the corresponding partial correlations.

	V1	V2	V3	V4	V5	V6
V1	-	0.028	0.08	0.076	0.045	0.117
V2	-	-	0.216	0.129	0.065	0.084
V3	-	-	-	0.633	0.063	0.080
V4	-	-	-	-	0.068	0.103
V5	-	-	-	-	-	0.049
V6	-	-	-	-	-	-

Table 2: Lower bound of the CI obtained with the bootstrap test for each zero-order correlation and corresponding partial correlation in Table 1.

	V1	V2	V3	V4	V5	V6
V1	1	-0.001	-0.406	-0.039	-0.257	-0.185
V2	0.004	1	0.215	0.034	0.233	0.310
V3	-0.520	-0.228	1	0.256	0.427	0.408
V4	0.480	0.254	0.820	1	0.780	-0.652
V5	-0.497	-0.115	-0.646	0.919	1	-0.156
V6	0.424	0.349	0.864	-0.950	0.816	1

Table 3: The upper triangle of the matrix represents a random correlation matrix. The correlations are randomly drawn from a uniform distribution that takes both positive and negative values. The lower triangle of the matrix represents the corresponding partial correlations. Each zero-order correlation with a gray background corresponds to a partial correlation with a gray background and the combination refers to a pair for which the bootstrap test was significant (the CI obtained with the bootstrap test did not include one).

	V1	V2	V3	V4	V5	V6
V1	-	0.071	0.557	3.159	0.847	2.439
V2	-	-	0.986	0.339	0.14	0.089
V3	-	-	-	2.396	2.761	1.99
V4	-	-	-	-	1.116	1.436
V5	-	-	-	-	-	4.793
V6	-	-	-	-	-	-

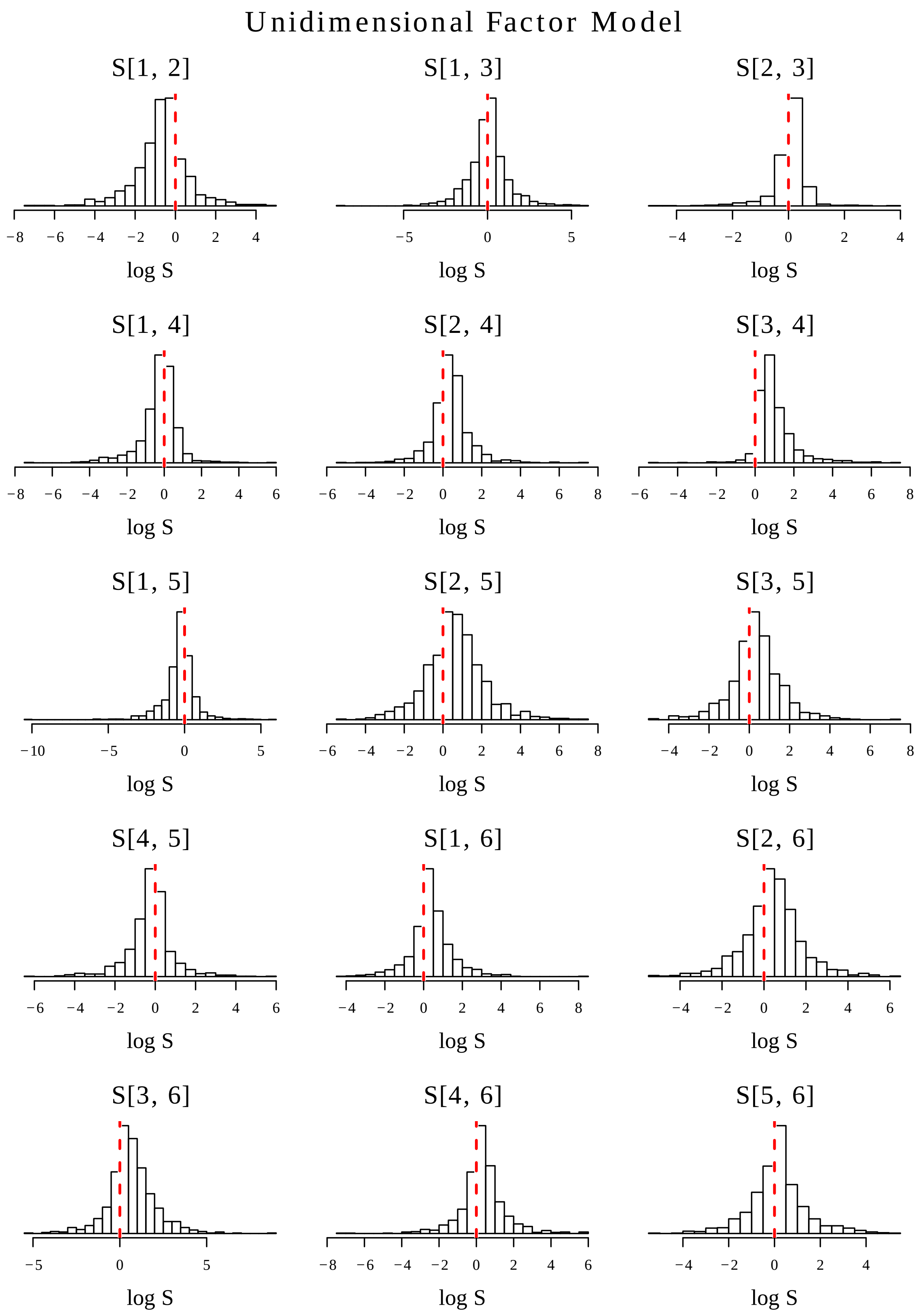
Table 4: Lower bound of the CI obtained with the bootstrap test for each zero-order correlation and corresponding partial correlation in Table 3. The values that have a gray background refer to a combination of a zero-order correlation and partial correlation for which the CI obtained with the bootstrap test does not include one.

		N									
model	# variables	100	200	300	400	500	600	700	800	900	1000
UFM	5	0.3	0.7	0.9	0.7	0.7	0.7	0.6	0.6	0.5	0.6
	10	2.8	2.4	2.9	1.8	2.7	1.9	2.1	2.4	2	1.7
	15	2.7	3.6	4.5	5.2	4.1	4.3	4.4	4.4	4.2	4.3
delta = 0.2	5	19.6	28.1	36.8	42.3	45.1	49.2	51.3	53	53.3	54.8
	10	25.9	42.4	51.9	55.2	53.9	57.9	62.4	64.1	62.6	64.5
	15	25.4	40.2	51.1	53.2	59.1	59.0	60.4	61.4	59.3	59.3
delta = 0.4	5	50.0	64.9	67.1	70.2	70.1	71.2	75.0	76.6	73.7	75.2
	10	51.7	62.1	65.7	68.2	69.5	70.0	69.3	69.6	70.9	72.3
	15	42.4	55.7	61.5	60.2	63.8	64.2	65.1	67.8	67.4	65.6
delta = 0.6	5	70.0	78.1	80.1	79.0	81.7	83.1	85	83.1	85	87.1
	10	55.5	66.2	73.1	73.8	75.2	74.6	76.5	76.9	74.1	77.0
	15	46.3	61.6	64.6	65.9	69.0	69.8	71.3	74.0	73.0	74.8
switch sign	5	62.4	73.9	83.7	85.1	87.8	89.8	91.6	91.9	93.5	94.5
	10	54.1	69.6	77.8	84.1	85.9	90.2	90.4	90.5	93.2	94.3
	15	48.1	67.7	75.5	81.5	83.5	87.4	88.3	91.1	92.0	92.2
random correlations	5	97.8	99.0	99.3	99.7	99.8	99.8	99.9	99.8	99.9	99.9
	10	100	100	100	100	100	100	100	100	100	100
	15	100	100	100	100	100	100	100	100	100	100

Table 5: Percentage of cases in which the bootstrap test rejects the UFM. The results in this table stem from simulations in which either a UFM was the data-generating model (i.e., percentages reflect the false positive rate) or an alternative model was the data-generating model (i.e., percentages reflect the power of the test). For this set of simulations all models resulted in solely positive correlations except the alternative model in which one of the correlations was switched from a positive to a negative sign.

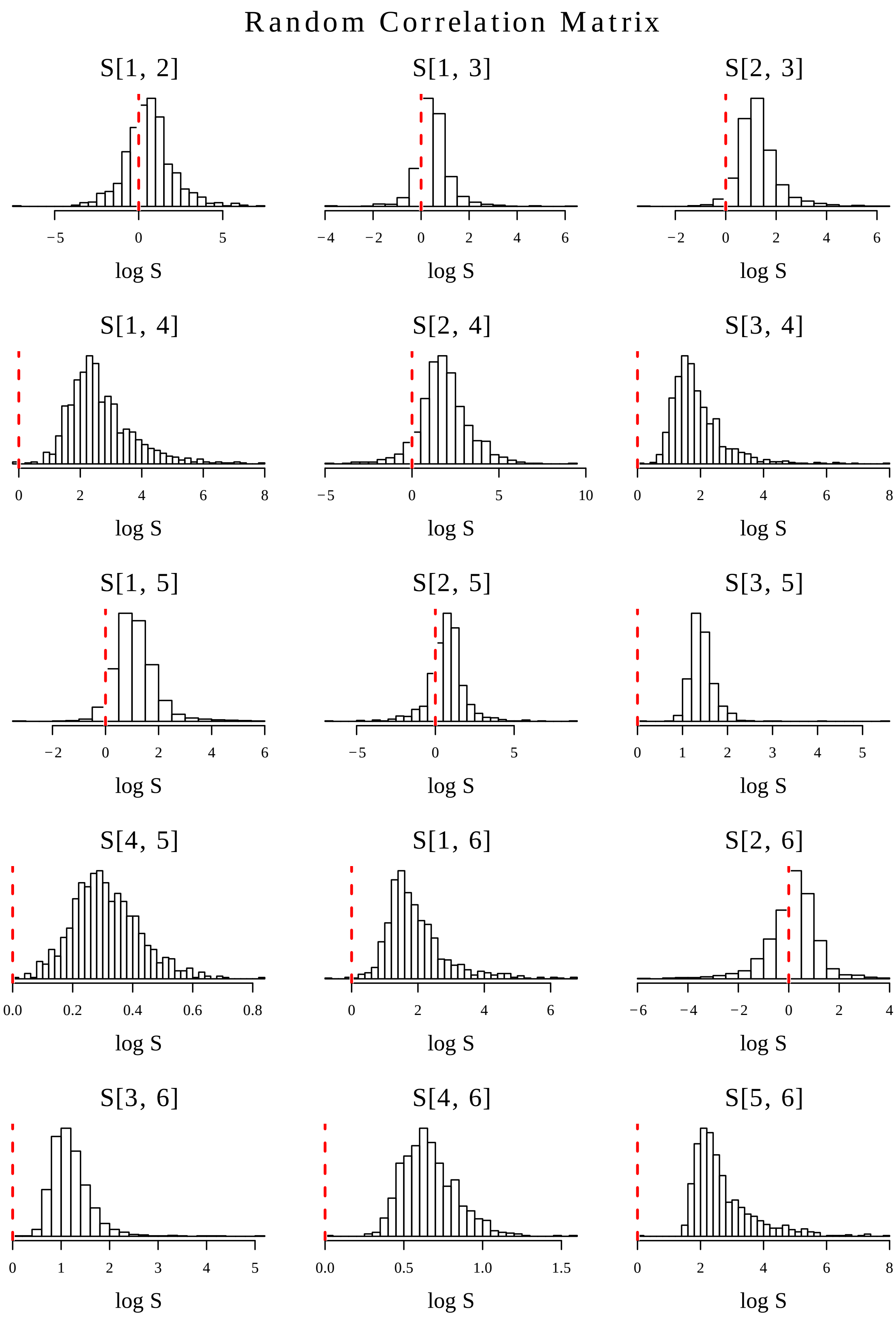
		N									
model	# variables	100	200	300	400	500	600	700	800	900	1000
UFM	5	0.5	0.9	0.5	0.1	0.5	0.4	0.9	0.6	0.1	0.6
	10	2.2	1.5	1.9	2	1.9	1.9	1.5	1.2	2.3	1.7
	15	2.8	3.5	3.4	3.2	3.7	4.4	3.3	4.6	3.7	3.5
delta = 0.2	5	15.4	27.3	33.3	36.1	40.2	43.1	47.9	45.3	46.2	48.7
	10	19.1	32.1	42.6	46.8	51.7	54.7	53.9	55.7	54.9	58.5
	15	22.5	33.8	43.5	49.4	50.0	54.6	57.9	51.4	57.4	57.6
delta = 0.4	5	46.1	55.1	63.5	66.9	67.5	74.0	72.8	71.1	74.9	73.8
	10	43.9	54.0	61.0	61.8	65.3	69.0	67.8	66.5	68.6	71.5
	15	37.0	50.2	55.3	56.3	61.0	62.4	63.8	65.8	65.6	65.6
delta = 0.6	5	61.4	70.4	76.2	78.4	79.9	82.0	82.5	83.0	82.7	84.4
	10	49.6	65.8	67.1	70.9	72.4	71.0	74.9	74.4	76.0	77.1
	15	42.6	54.9	62.6	65.3	66.4	66.9	68.1	72.4	70.3	71.8
switch sign	5	46.5	56.1	64.3	67.4	70.8	73.7	73.2	75.7	75.9	75.2
	10	37.9	52.2	62.1	62.2	67.8	68.0	69.3	75.1	73.4	71.5
	15	33.7	51.6	55.4	61.1	66.5	66.0	72.2	70.6	73.0	74.8
random correlations	5	97.8	99	99.3	99.7	99.8	99.8	99.9	99.8	99.9	99.9
	10	96.8	99.1	99.4	99.4	99.7	99.8	100	100	99.8	100
	15	100	100	100	100	100	100	100	100	100	100

Table 6: Percentage of cases in which the bootstrap test rejects the UFM. The results in this table stem from simulations in which either a UFM was the data-generating model (i.e., percentages reflect the false positive rate) or an alternative model was the data-generating model (i.e., percentages reflect the power of the test). For this set of simulations all models resulted in correlation matrices with both positive and negative correlations.


Unidimensional Factor Model
S[1, 2]
S[1, 3]
S[2, 3]
S[1, 4]
S[2, 4]
S[3, 4]
S[1, 5]
S[2, 5]
S[3, 5]
S[4, 5]
S[1, 6]
S[2, 6]
S[3, 6]
S[4, 6]
S[5, 6]
log S

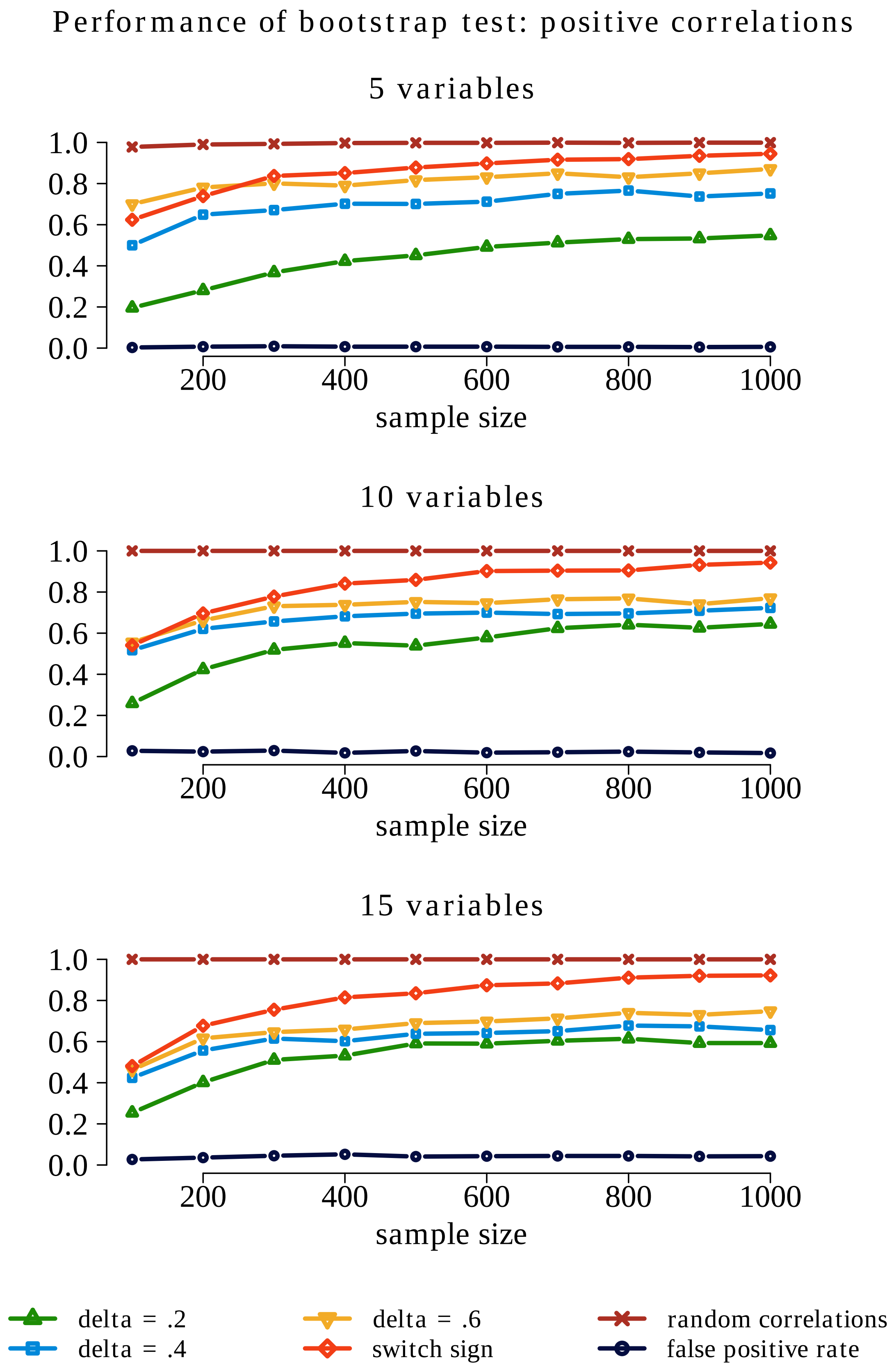
:

Figure 1 *(previous page)*: Bootstrap results for the simulated data example. The results are based on a dataset with 6 variables and 60 observations. The horizontal axis represents $\log(|2\rho_{y_1 y_2 \cdot Z}/\rho_{y_1 y_2} - 1|)$. The vertical dashed red line indicates the point where $\log(|2\rho_{y_1 y_2 \cdot Z}/\rho_{y_1 y_2} - 1|) = 0$ (i.e., $|2\rho_{y_1 y_2 \cdot Z}/\rho_{y_1 y_2} - 1| = 1$). Variables for which the CI of the bootstrapped values only includes values greater than 1 provide evidence against the UFM.


Random Correlation Matrix
S[1, 2]
S[1, 3]
S[2, 3]
S[1, 4]
S[2, 4]
S[3, 4]
S[1, 5]
S[2, 5]
S[3, 5]
S[4, 5]
S[1, 6]
S[2, 6]
S[3, 6]
S[4, 6]
S[5, 6]
log S

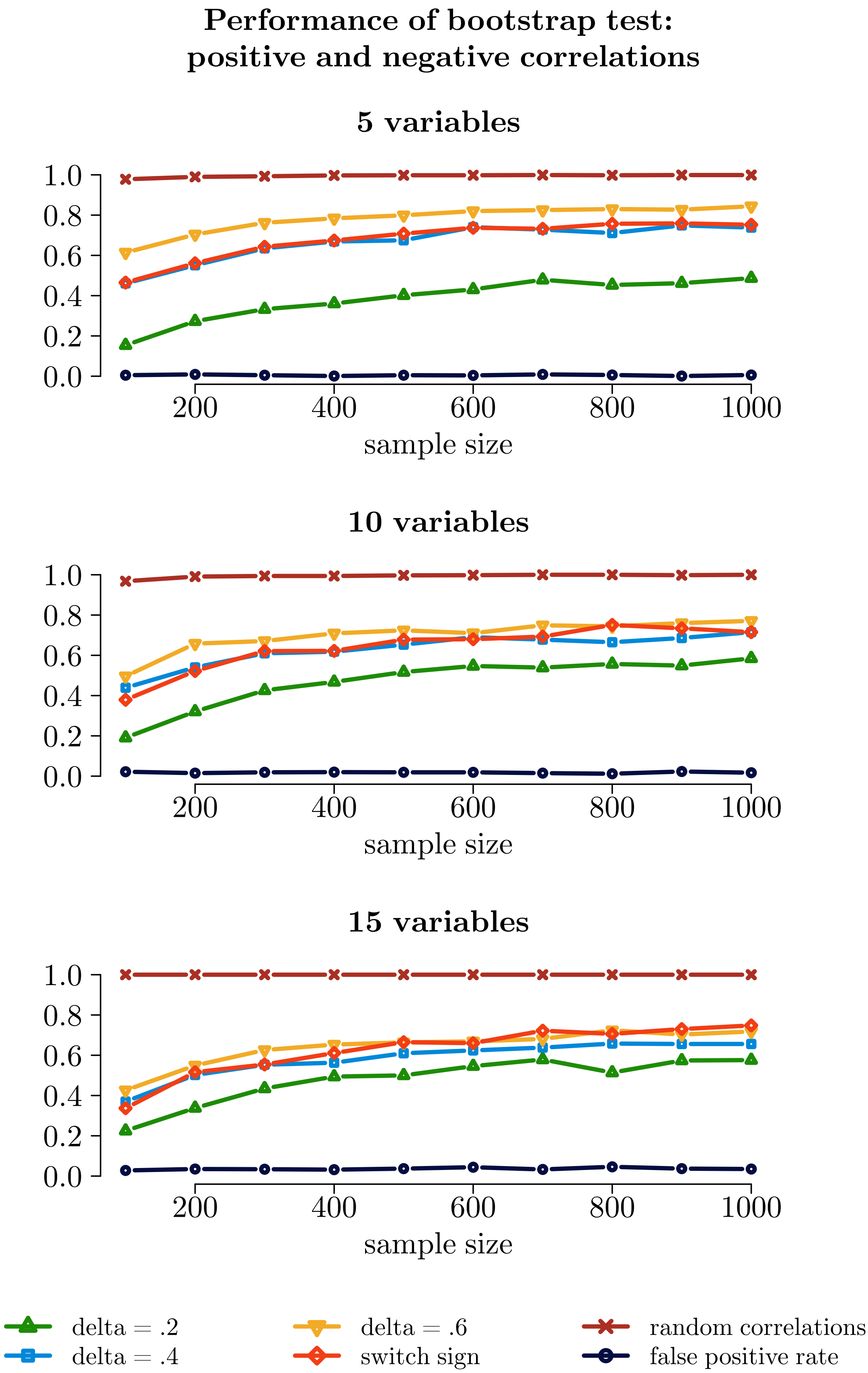
:

Figure 2 *(previous page)*: Bootstrap results when the data were generated from a random correlation matrix with both positive and negative correlations. The results are based on a dataset with 6 variables and 60 observations. The horizontal axis represents $\log(|2\rho_{y_1y_2 \cdot Z}/\rho_{y_1y_2} - 1|)$. The vertical dashed red line indicates the point where $\log(|2\rho_{y_1y_2 \cdot Z}/\rho_{y_1y_2} - 1|) = 0$ (i.e., $|2\rho_{y_1y_2 \cdot Z}/\rho_{y_1y_2} - 1| = 1$). Variables for which the CI of the bootstrapped values only includes values greater than 1 provide evidence against the unidimensional factor model.


Performance of bootstrap test: positive correlations
5 variables
1.0
0.8
0.6
0.4
0.2
0.0
200
400
600
800
1000
sample size
10 variables
1.0
0.8
0.6
0.4
0.2
0.0
200
400
600
800
1000
sample size
15 variables
1.0
0.8
0.6
0.4
0.2
0.0
200
400
600
800
1000
sample size
delta = .2
delta = .4
delta = .6
switch sign
random correlations
false positive rate

:

Figure 3 *(previous page)*: Performance of the bootstraptest on simulated data with only positive correlations.


Performance of bootstrap test:
positive and negative correlations
5 variables
10 variables
15 variables
1.0
0.8
0.6
0.4
0.2
0.0
200
400
600
800
1000
sample size
delta = .2
delta = .4
delta = .6
switch sign
random correlations
false positive rate

:

Figure 4 *(previous page)*: Performance of the bootstraptest on simulated data with both positive and negative correlations.