# Low level of covariance simulation

*Xuelong Wang*

*2019-08-23*

## Contents



# 1 Motivation

Based on the investigation of the covariance matrix patter of the PCBs from 1999 - 2004, we found some patterns that appears among the covariance matrix of PCBs from differert years. Besides, we also use the historical data to estimate the covariance matrix, then use the sample covariance matrix to decorrelate the PCBs for each year. For most years, the decorrelation procedure can reduce the correlation, e.g 1999-2001, but there are some years which the correlations still are high after decorrelation, e.g 2005-2006.

Since we are trying to borrow information of historical dataset, the decorrelated data will probably not be perfert uncorrelted. In other words, we will end up with data with low correlations among their coveriates. So the goal of this report is to see if the GCTA and EigenPrism method could work well under the low correlation setup.

# 2 setup

## 2.1 Sample covariance matrix

I use two different sample covariane matrices: 1999-2001 and 2007 -2008.

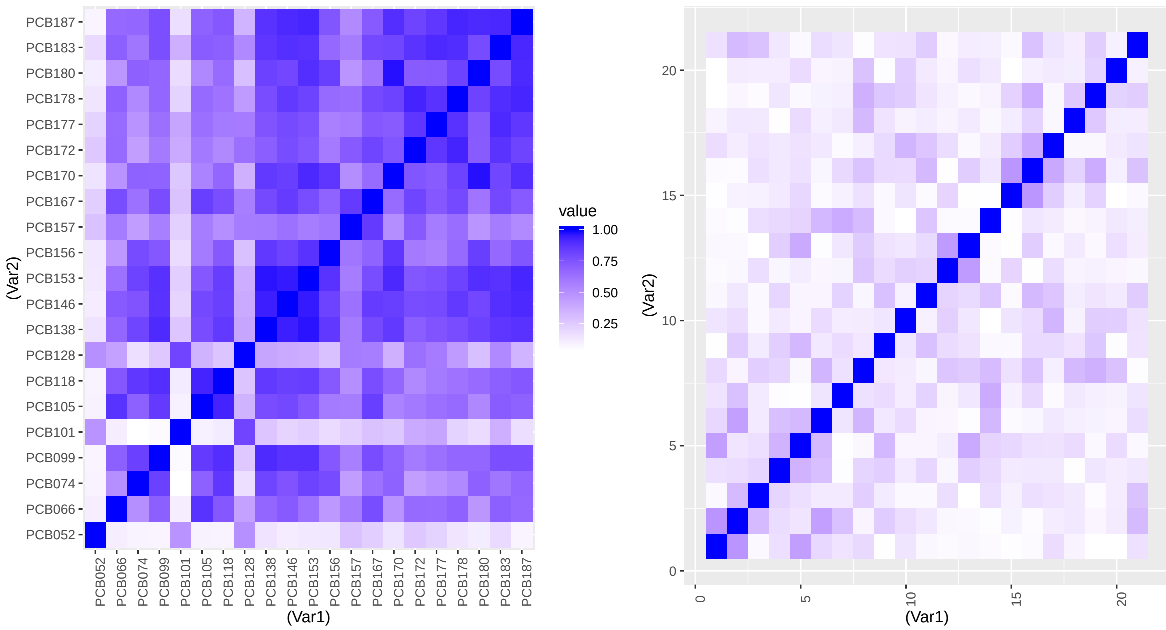



Figure 1: 1999-2000

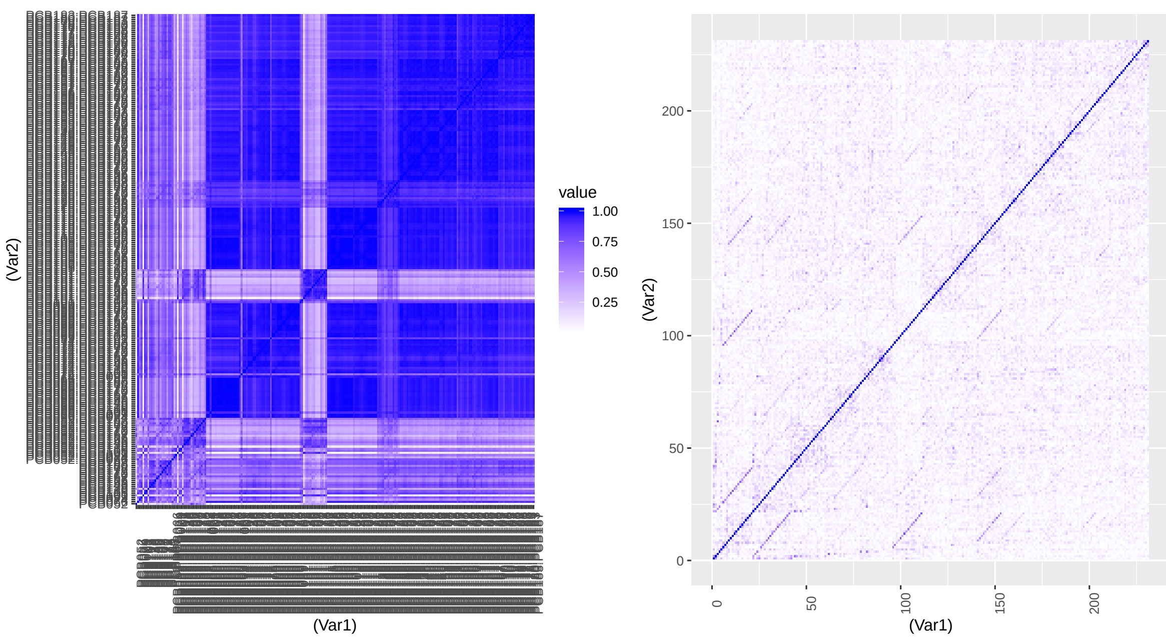
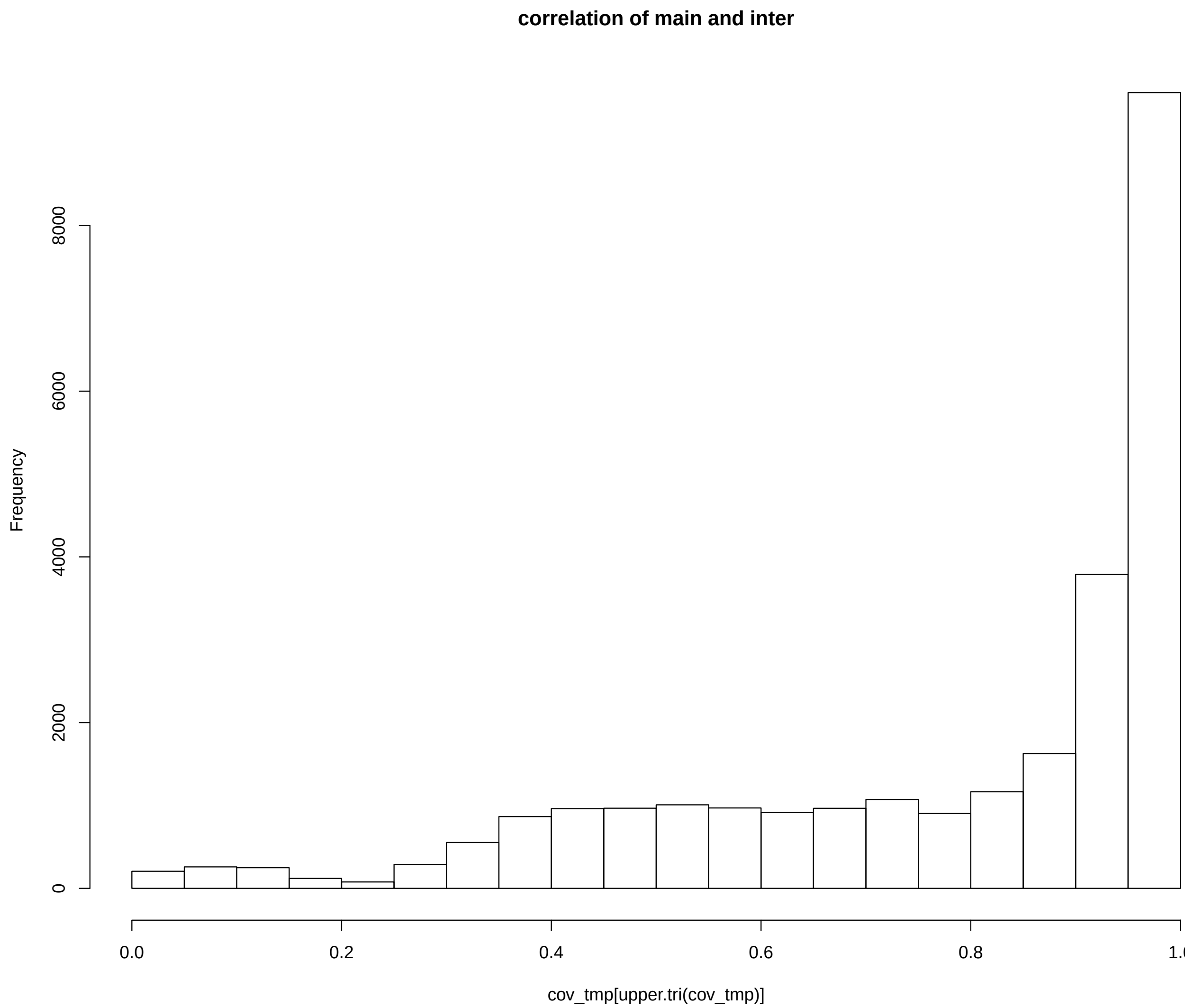
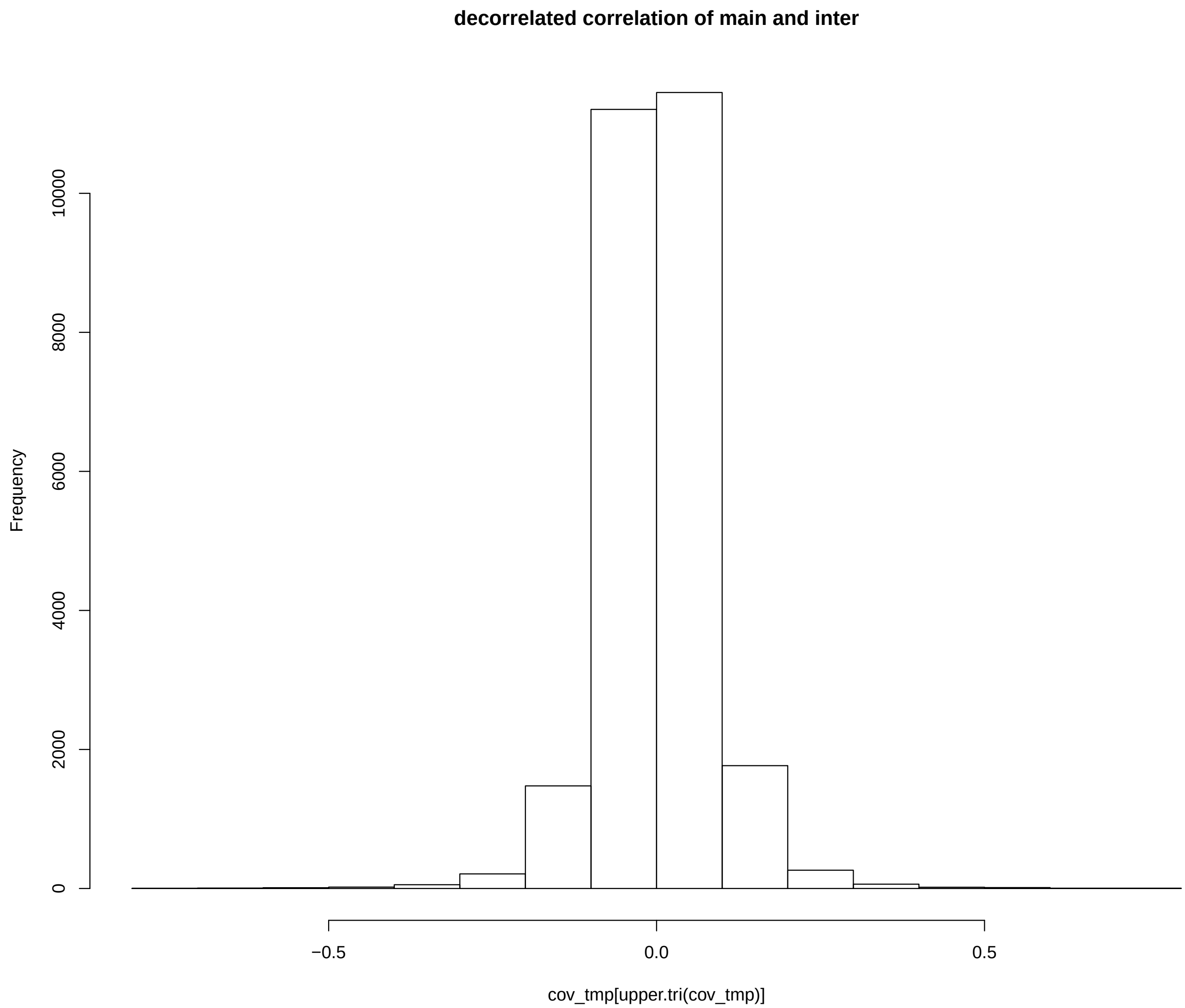



Figure 2: Combined main and interaction 1999-2000


correlation of main and inter
Frequency
0
2000
4000
6000
8000
0.0
0.2
0.4
0.6
0.8
1.0
cov_tmp[upper.tri(cov_tmp)]


decorrelated correlation of main and inter
Frequency
0
2000
4000
6000
8000
10000
−0.5
0.0
0.5
cov_tmp[upper.tri(cov_tmp)]

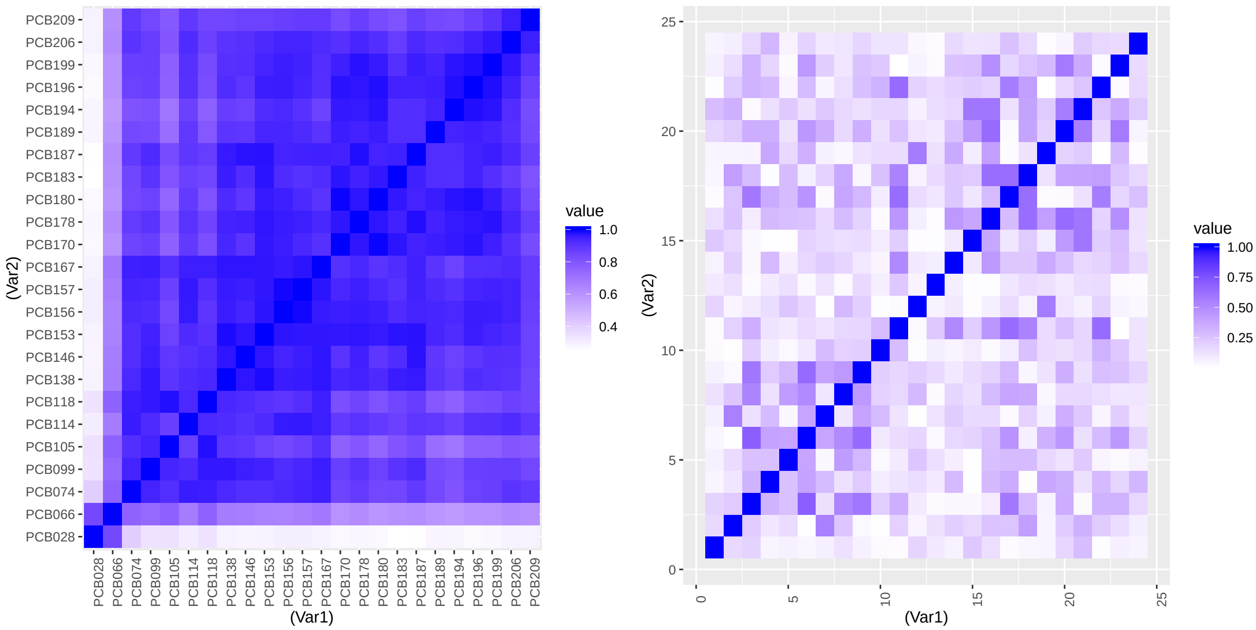



Figure 3: 2005-2006

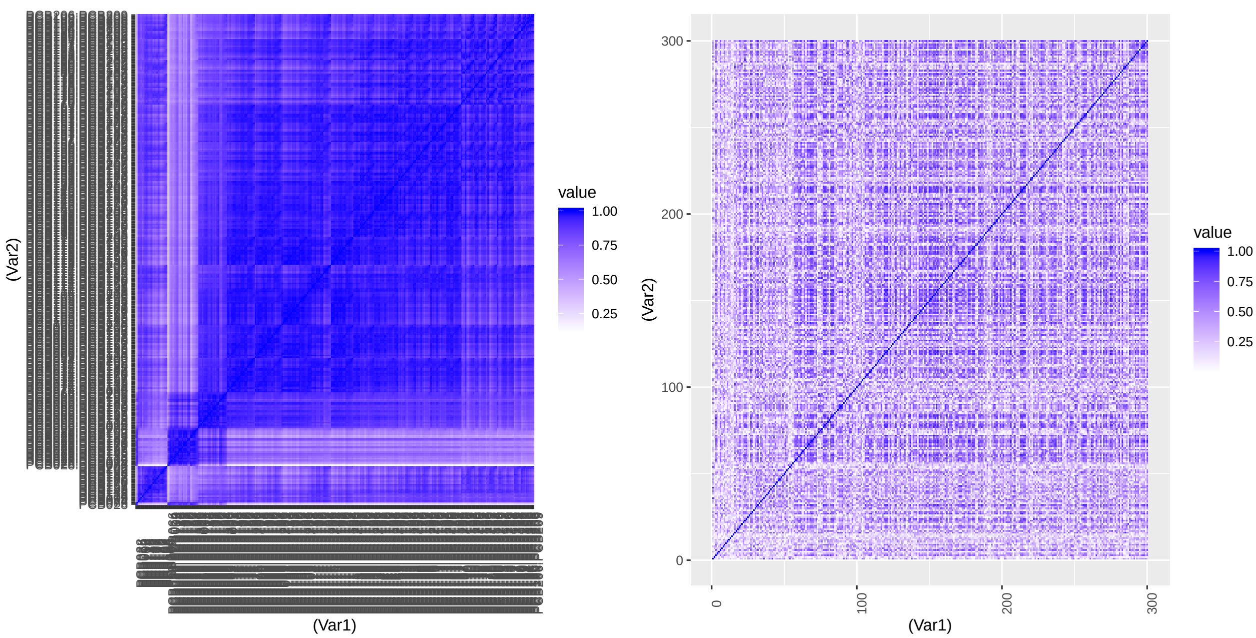
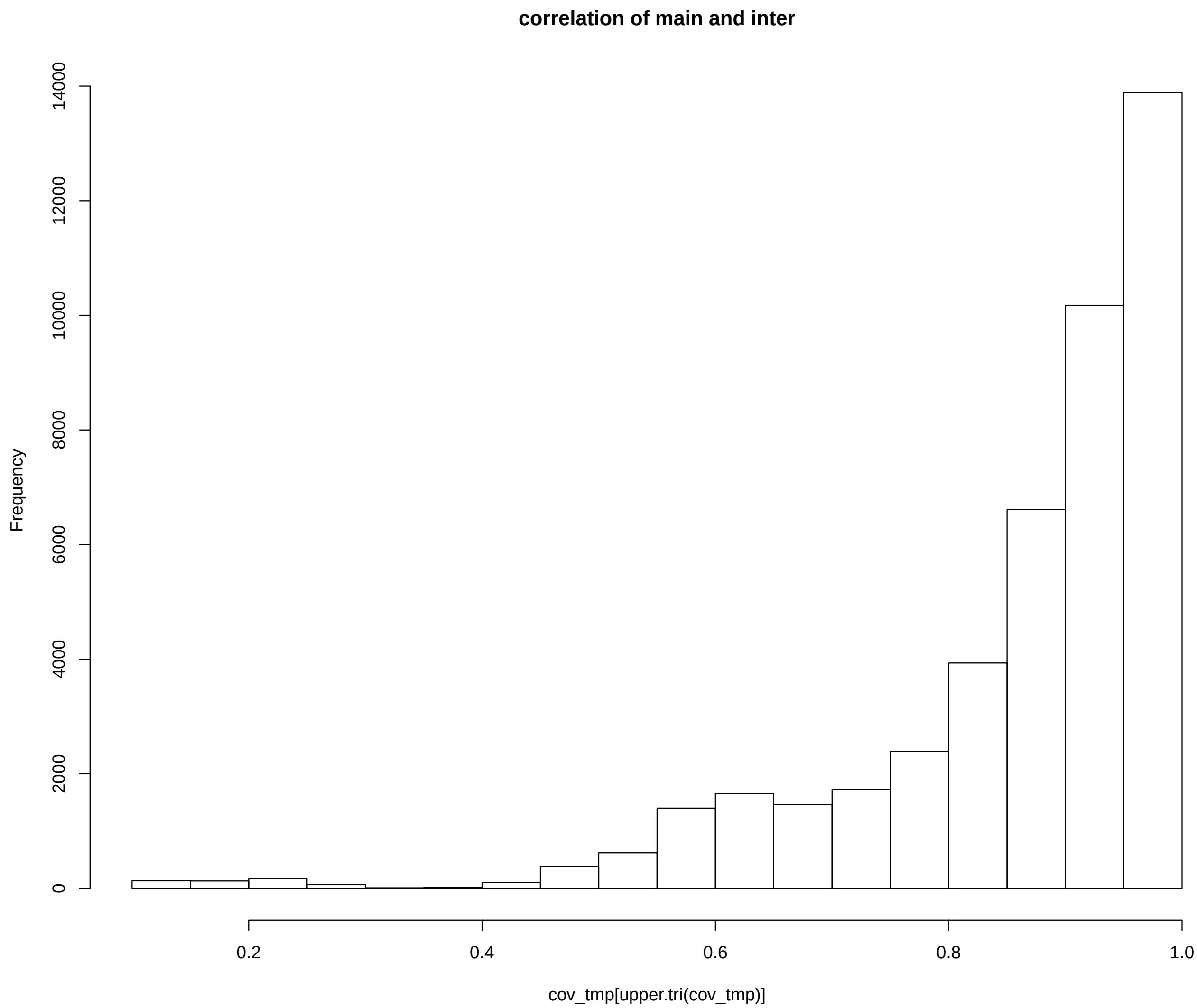
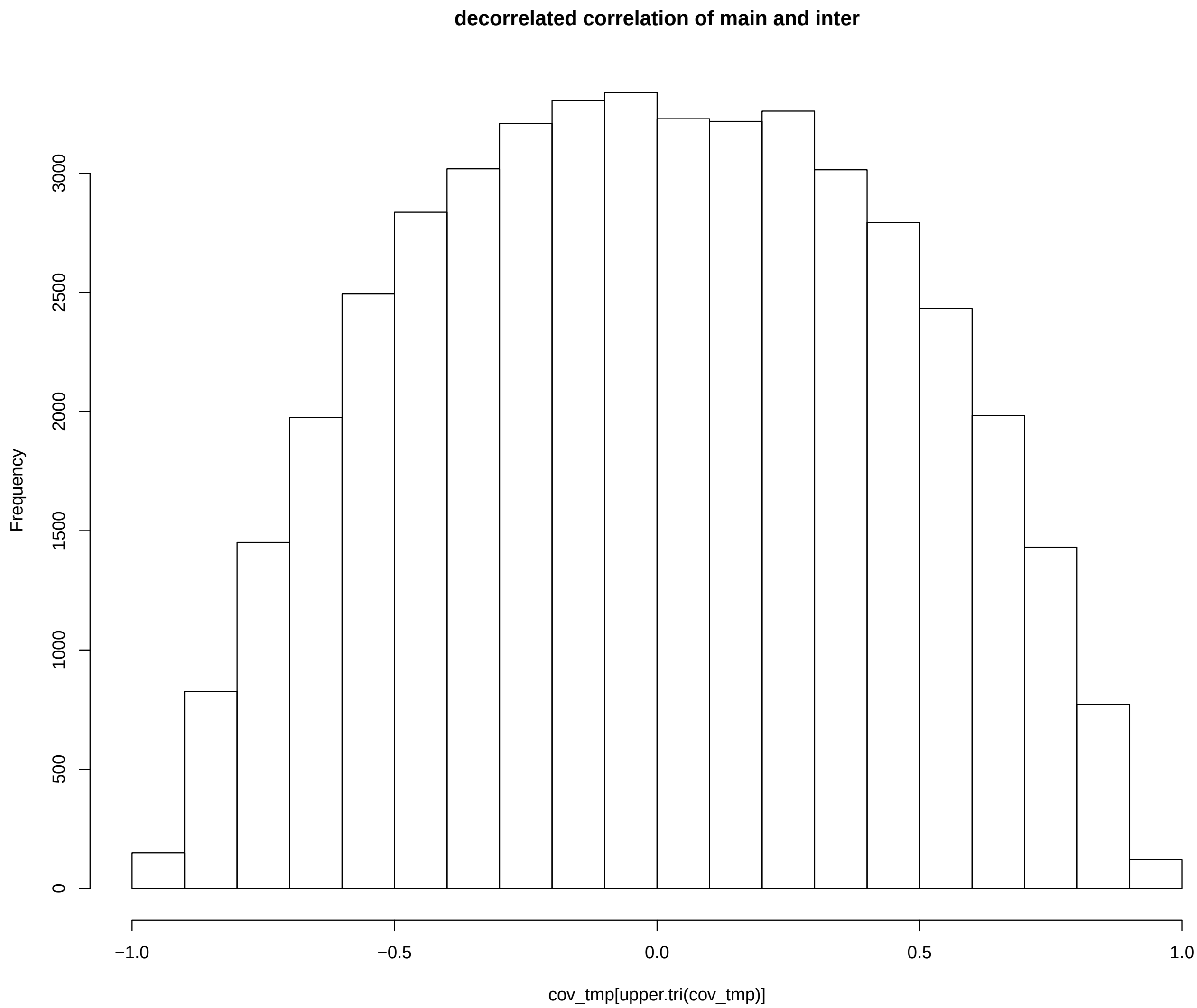



Figure 4: Combined main and interaction 2005-2006


correlation of main and inter
Frequency
0
2000
4000
6000
8000
10000
12000
14000
0.2
0.4
0.6
0.8
1.0
cov_tmp[upper.tri(cov_tmp)]


decorrelated correlation of main and inter
Frequency
0
500
1000
1500
2000
2500
3000
−1.0
−0.5
0.0
0.5
1.0
cov_tmp[upper.tri(cov_tmp)]

## 2.2 Simulation setup

1. p = 231 for 1999 or 300 for 2005
2. $n \in \{100, 150, p\}$
3. the simulated covariates will follow Normal or Chi with df = 1
4. Variance estimation method will be GCTA and EigenPrism method
5. Iteration time is 100

# 3 Result

## 3.1 1999

```
      n  MSE est_var est_mean     method target var_main_effect decor
 1: 100 19.6    18.9     10.9 EigenPrism   main              10 FALSE
 2: 150  8.7     8.7     10.2 EigenPrism   main              10 FALSE
 3: 231  5.5     5.5     10.1 EigenPrism   main              10 FALSE
 4: 100 21.3    21.4     10.3       GCTA   main              10 FALSE
 5: 150  9.3     9.4     10.1       GCTA   main              10 FALSE
 6: 231  5.0     5.1     10.0       GCTA   main              10 FALSE
 7: 100 20.6    20.6     10.4 EigenPrism   main              10 FALSE
 8: 150 10.4    10.4      9.7 EigenPrism   main              10 FALSE
 9: 231  5.3     5.2     10.4 EigenPrism   main              10 FALSE
10: 100 20.7    21.0     10.0       GCTA   main              10 FALSE
11: 150 10.9    10.6      9.4       GCTA   main              10 FALSE
12: 231  4.8     4.7     10.4       GCTA   main              10 FALSE
    x_dist
 1:    chi
 2:    chi
 3:    chi
 4:    chi
 5:    chi
 6:    chi
 7: normal
 8: normal
 9: normal
10: normal
11: normal
12: normal
```

## 3.2 2005

```
      n  MSE est_var est_mean     method target var_main_effect decor
 1: 100 16.4    15.8     10.9 EigenPrism   main              10 FALSE
 2: 150 13.3    11.6     11.3 EigenPrism   main              10 FALSE
 3: 300  4.5     4.6      9.9 EigenPrism   main              10 FALSE
 4: 100 12.3    11.2     11.1       GCTA   main              10 FALSE
 5: 150 11.0     7.7     11.8       GCTA   main              10 FALSE
```

```
 6: 300  3.7     3.0     10.9       GCTA   main              10 FALSE
 7: 100 15.1    13.6     11.3 EigenPrism   main              10 FALSE
 8: 150 11.4    10.6     11.0 EigenPrism   main              10 FALSE
 9: 300  6.7     6.7     10.3 EigenPrism   main              10 FALSE
10: 100 12.1    10.2     11.4       GCTA   main              10 FALSE
11: 150  9.4     7.2     11.5       GCTA   main              10 FALSE
12: 300  4.4     3.4     11.0       GCTA   main              10 FALSE
    x_dist
 1:    chi
 2:    chi
 3:    chi
 4:    chi
 5:    chi
 6:    chi
 7: normal
 8: normal
 9: normal
10: normal
11: normal
12: normal
```