

GEE Comparison

Jon Ong

2025-03-16

Simulation Setup

	iter	n	p	QELR Param
Sim1	500	300	3	V1
Sim2	500	300	3	V2
Sim3	500	300	5	V3
Sim4	500	300	10	V4
Sim5	500	300	12	V5

Remarks:

1. I've excluded the “unstructured” correlation structure from simulation.
2. In terms of gee parameters:
 1. Maxiter: Manually increased maxiter to a large number (eg. 200), when working correlation is not independent.
 2. Scale.fixed: Sometimes solutions would converge when scale.fixed = TRUE (default = FALSE). Currently set as FALSE
3. GEE package: only used the sandwich estimator's result (Robust variance)

```
allout$V1 %>% round(.,2)
```

```
##      [,1] [,2] [,3]
## [1,] -1.2 -0.2  0.2
## [2,] -0.2 -0.1 -0.2
## [3,]  0.2 -0.2  1.0
```

```
allout$V2 %>% round(.,2)
```

```
##      [,1] [,2] [,3]
## [1,] -1.2  0.2  0.2
## [2,]  0.2 -0.1  0.2
## [3,]  0.2  0.2  1.0
```

```
allout$V3 %>% round(.,2)
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,] -1.5 -0.20  0.2 -0.20  0.2
## [2,] -0.2 -0.75 -0.2  0.20 -0.2
## [3,]  0.2 -0.20  0.0 -0.20  0.2
## [4,] -0.2  0.20 -0.2  0.75 -0.2
## [5,]  0.2 -0.20  0.2 -0.20  1.5
```

```
allout$V4 %>% round(.,2)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,] -1.5 -0.20  0.20 -0.2  0.20 -0.20  0.2 -0.20  0.20 -0.2
## [2,] -0.2 -1.17 -0.20  0.2 -0.20  0.20 -0.2  0.20 -0.20  0.2
## [3,]  0.2 -0.20 -0.83 -0.2  0.20 -0.20  0.2 -0.20  0.20 -0.2
## [4,] -0.2  0.20 -0.20 -0.5 -0.20  0.20 -0.2  0.20 -0.20  0.2
## [5,]  0.2 -0.20  0.20 -0.2 -0.17 -0.20  0.2 -0.20  0.20 -0.2
## [6,] -0.2  0.20 -0.20  0.2 -0.20  0.17 -0.2  0.20 -0.20  0.2
## [7,]  0.2 -0.20  0.20 -0.2  0.20 -0.20  0.5 -0.20  0.20 -0.2
## [8,] -0.2  0.20 -0.20  0.2 -0.20  0.20 -0.2  0.83 -0.20  0.2
## [9,]  0.2 -0.20  0.20 -0.2  0.20 -0.20  0.2 -0.20  1.17 -0.2
## [10,] -0.2  0.20 -0.20  0.2 -0.20  0.20 -0.2  0.20 -0.20  1.5
```

Differences Between GEE and GEEMPACK: Do both packages generate the same estimates?

Short answer: YES for independence correlation.

GEE stops and returns an error message when solution diverges, however GEEMPACK would still produce (erroneous, some extremely large values) estimates.

The below analysis answers a few questions to address the comparison/ differences:

1. Out of 500 simulations, how many converged solutions do we get?

```
rbind(sim1= sapply(allout$sim1$convergecase, sum),
      sim2= sapply(allout$sim2$convergecase, sum),
      sim3= sapply(allout$sim3$convergecase, sum)) %>% t()
```

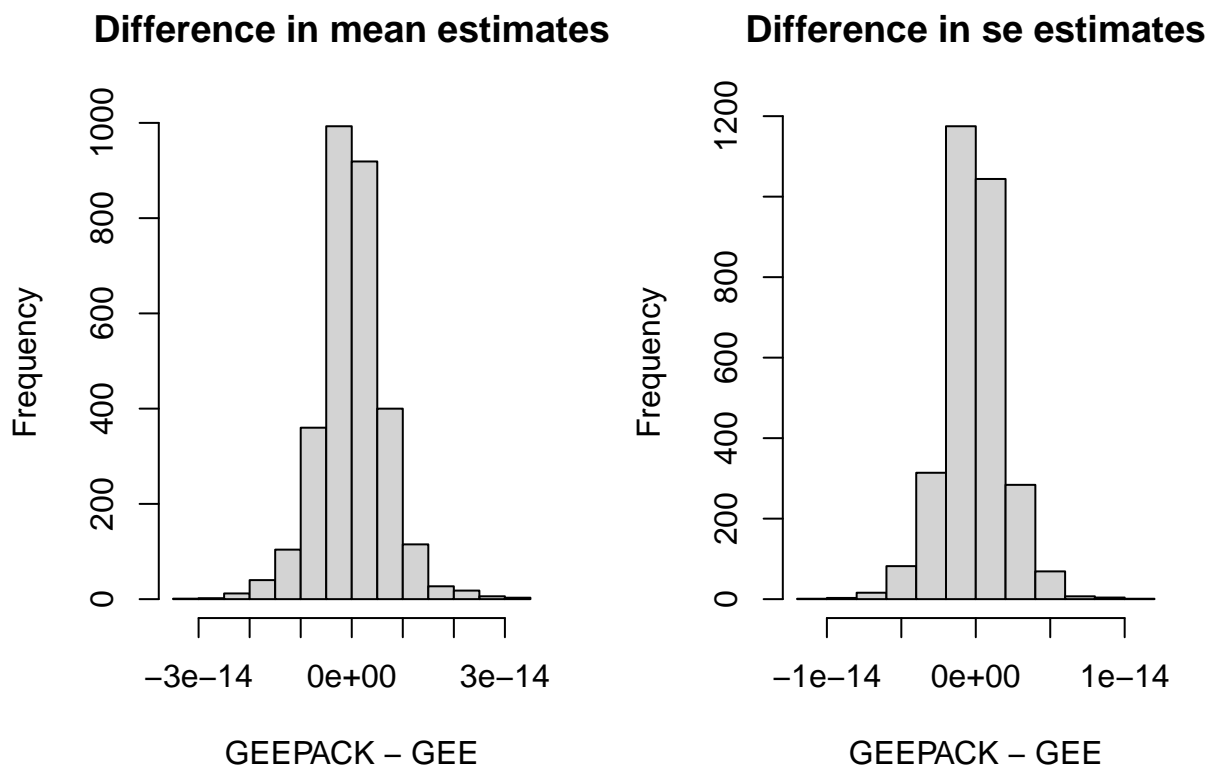
```
##      sim1 sim2 sim3
## logit      500  500  500
## geepack_ind 500  500  500
## geepack_exc 497  500  500
## geepack_ar1 500  500  500
## gee_ind     500  500  500
## gee_exc      72  490   93
## gee_ar1       1    0    0
## mle         500  500  500
```

1. Distribution of the differences between GEE and GEEMPACK estimates (among the converged cases)?

Sim1

1. Independence correlation shows negligible differences.
2. Exchangeable correlation shows noticeable differences. This refers to the 72 converged cases from “GEE”. Could it be that GEEPACK and GEE yield different divergence/ convergence cases? Further deep dive can be done here.
3. AR-1: only 1 converged case here, no comment.

```
checkdiff(allout$sim1$OUT$geepack_ind, allout$sim1$OUT$gee_ind)
```

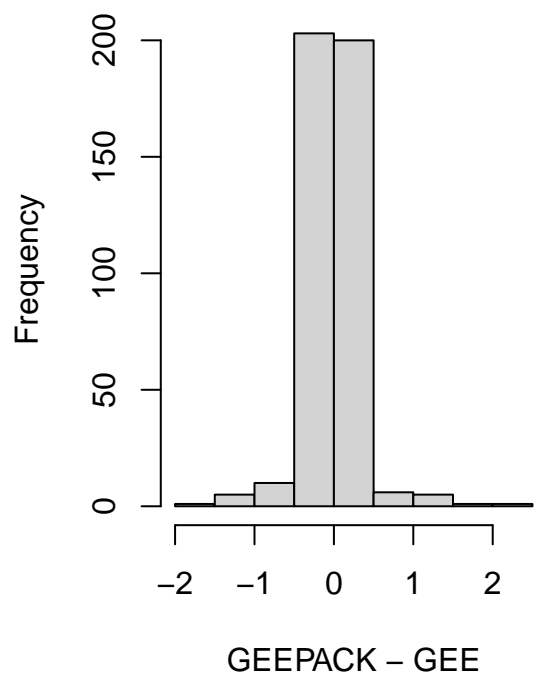


```
## number of convergence: 500 out of 500
```

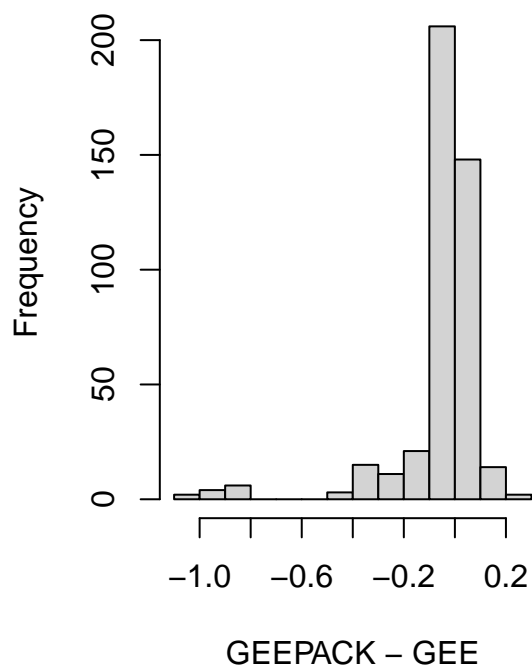
##	Estimate	Std. Error
## Min.	:-3.020e-14	Min. :-1.005e-14
## 1st Qu.	:-3.608e-15	1st Qu.:-1.221e-15
## Median	: 0.000e+00	Median :-1.110e-16
## Mean	: 8.142e-17	Mean :-1.129e-16
## 3rd Qu.	: 3.719e-15	3rd Qu.: 1.055e-15
## Max.	: 3.353e-14	Max. : 1.005e-14

```
checkdiff(allout$sim1$OUT$geepack_exc, allout$sim1$OUT$gee_exc)
```

Difference in mean estimates



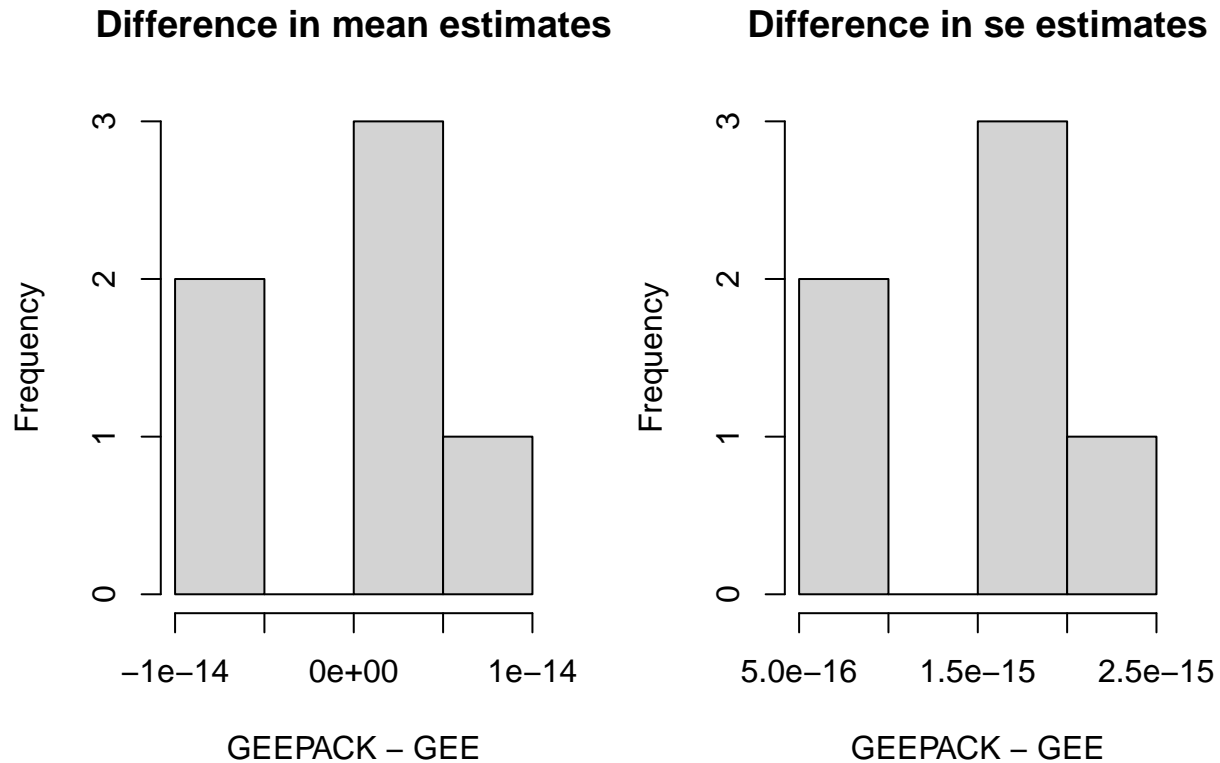
Difference in se estimates



number of convergence: 72 out of 500

##	Estimate	Std. Error
##	Min. : -1.6935958	Min. : -1.0862177
##	1st Qu.: -0.0118562	1st Qu.: -0.0268160
##	Median : -0.0000346	Median : -0.0006945
##	Mean : 0.0008746	Mean : -0.0491836
##	3rd Qu.: 0.0128033	3rd Qu.: 0.0028516
##	Max. : 2.0350040	Max. : 0.2490535

```
checkdiff(allout$sim1$OUT$geepack_ar1, allout$sim1$OUT$gee_ar1)
```



number of convergence: 1 out of 500

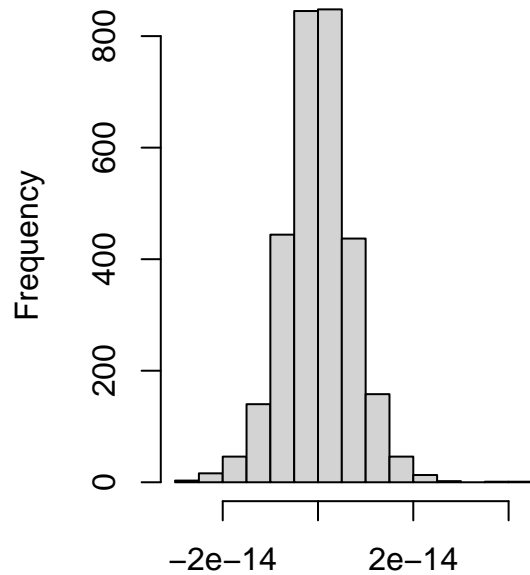
##	Estimate	Std. Error
## Min.	:-7.994e-15	Min. :8.604e-16
## 1st Qu.	:-5.621e-15	1st Qu.:1.152e-15
## Median	: 3.469e-16	Median :1.749e-15
## Mean	:-4.441e-16	Mean :1.624e-15
## 3rd Qu.	: 2.859e-15	3rd Qu.:1.929e-15
## Max.	: 8.549e-15	Max. :2.442e-15

Sim2

1. Independence correlation shows negligible differences.
2. Exchangeable correlation shows noticeable differences in both packages, using the 490 converged cases from GEE.
3. AR-1: 0 converged case here, no comparison/ comment.

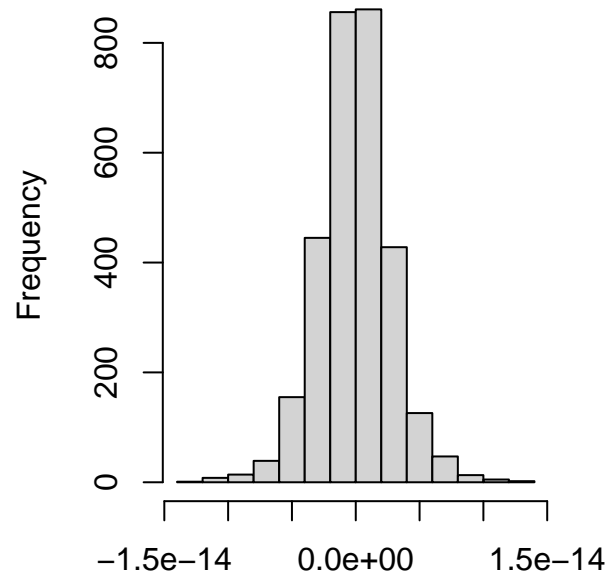
```
checkdiff(allout$sim2$OUT$geepack_ind, allout$sim2$OUT$gee_ind)
```

Difference in mean estimates



GEEPACK - GEE

Difference in se estimates



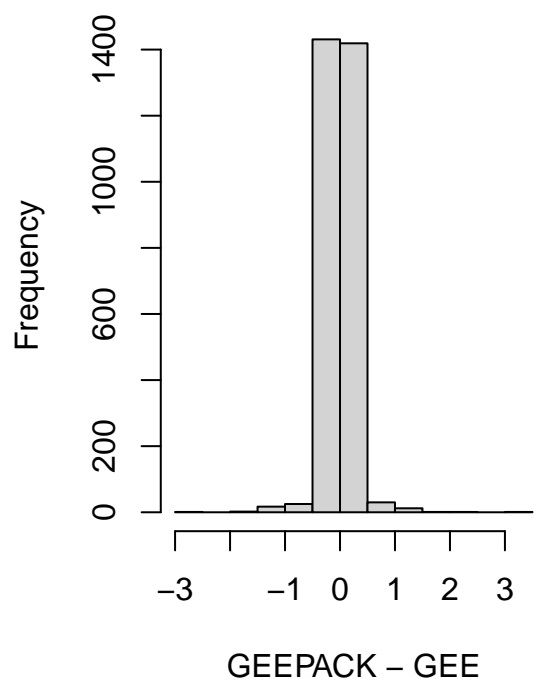
GEEPACK - GEE

```
## number of convergence: 500 out of 500
```

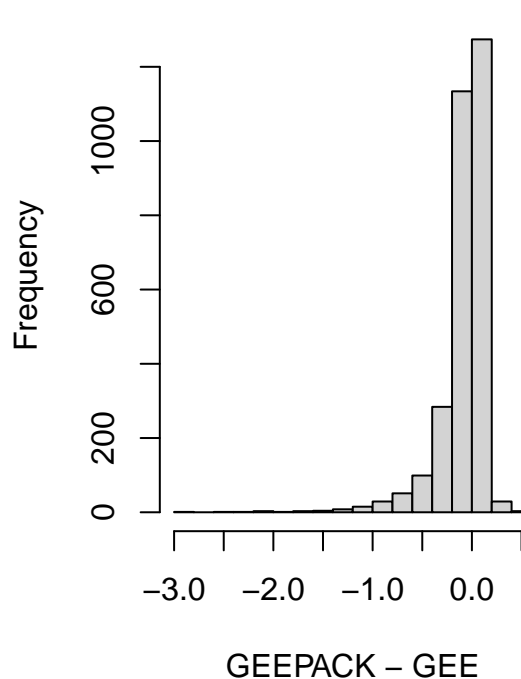
```
##      Estimate      Std. Error
## Min.   :-2.798e-14  Min.    :-1.227e-14
## 1st Qu.: -4.219e-15  1st Qu.: -1.776e-15
## Median :  5.551e-17  Median :  0.000e+00
## Mean    :  6.861e-17  Mean     :-5.108e-17
## 3rd Qu.:  4.365e-15  3rd Qu.:  1.665e-15
## Max.    :  4.063e-14  Max.     :  1.321e-14
```

```
checkdiff(allout$sim2$OUT$geepack_exc, allout$sim2$OUT$gee_exc)
```

Difference in mean estimates



Difference in se estimates



```
## number of convergence: 490 out of 500
```

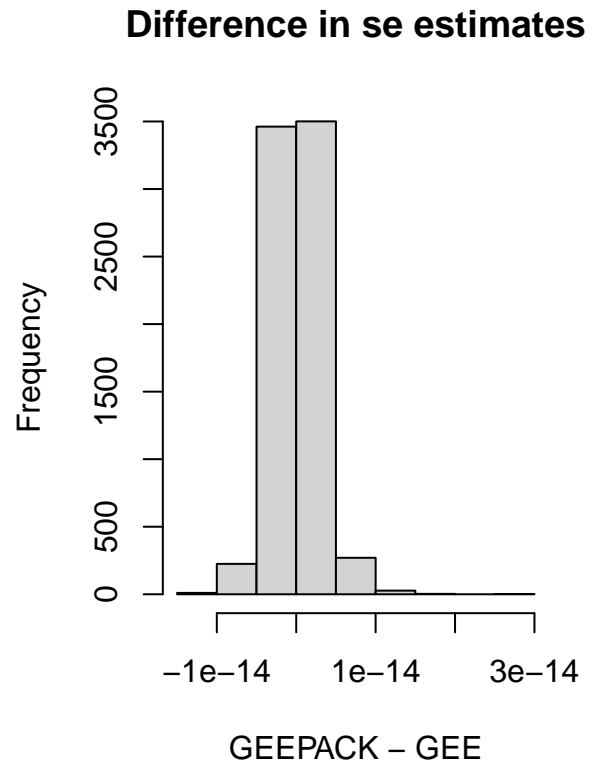
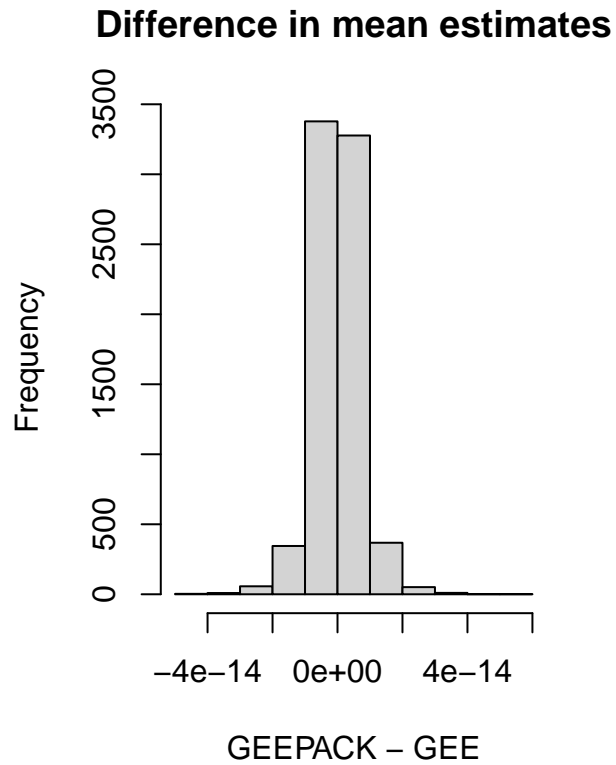
```
##      Estimate      Std. Error
## Min.    :-2.859892   Min.     :-2.8003126
## 1st Qu.: -0.005257   1st Qu.: -0.1093432
## Median :  0.000000   Median : -0.0002693
## Mean    :  0.000007   Mean      :-0.0877836
## 3rd Qu.:  0.005852   3rd Qu.:  0.0046210
## Max.    :  3.394594   Max.      :  0.4927737
```

```
#checkdiff(allout$sim2$OUT$geepack_ar1, allout$sim2$OUT$gee_ar1) # no convergence cases
```

Sim3

1. Independence correlation shows negligible differences.
2. Exchangeable correlation shows noticeable differences in both packages, using the 93 converged cases from GEE.
3. AR-1: 0 converged case here, no comparison/ comment.

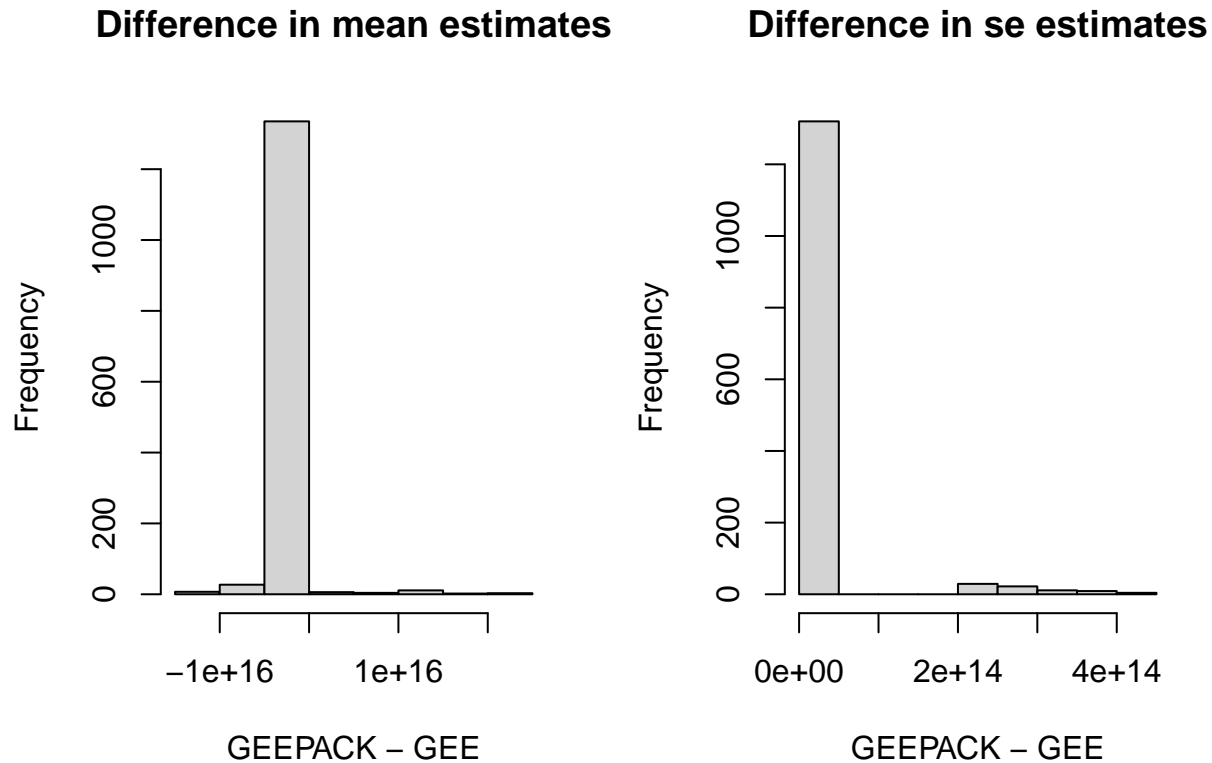
```
checkdiff(allout$sim3$OUT$geepack_ind, allout$sim3$OUT$gee_ind)
```



```
## number of convergence: 500 out of 500
```

```
##      Estimate      Std. Error
## Min.   :-4.818e-14  Min.    :-1.471e-14
## 1st Qu.: -3.109e-15  1st Qu.: -1.166e-15
## Median :  0.000e+00  Median :  5.551e-17
## Mean   :  2.910e-17  Mean    :  1.084e-16
## 3rd Qu.:  3.109e-15  3rd Qu.:  1.277e-15
## Max.   :  5.018e-14  Max.    :  2.653e-14
```

```
checkdiff(allout$sim3$OUT$geepack_exc, allout$sim3$OUT$gee_exc)
```

```
## number of convergence: 93 out of 500
```

```
##      Estimate      Std. Error
## Min.      :-1.497e+16  Min.      :0.000e+00
## 1st Qu.: 0.000e+00    1st Qu.:0.000e+00
## Median : 0.000e+00    Median :0.000e+00
## Mean      :-1.775e+13   Mean      :1.533e+13
## 3rd Qu.: 0.000e+00    3rd Qu.:0.000e+00
## Max.      : 2.108e+16   Max.      :4.378e+14
```

```
##checkdiff(allout$sim3$OUT$geepack_ar1, allout$sim3$OUT$gee_ar1) # no convergence cases
```

Differences across different working correlations: Do all correlation structures generate the same estimates?

This can be answered by checking the bias and relative s.e:

1. Independence correlation: already proved from previous simulation that it guarantees good estimates for parameters.
2. Exchangeable correlation: poor estimates based on the converged results from GEE.
3. AR1 correlation: almost no convergence, hence no comment.

Sim1

```
##      truth true_se estmean_gee_ind bias_gee_ind se_gee_ind re_gee_ind
## [1,]  -1.2  0.2972      -1.2105     -0.0105     0.2984     1.0039
```

##	[2,]	-0.4	0.2980	-0.4087	-0.0087	0.2861	0.9602
##	[3,]	0.4	0.3058	0.4064	0.0064	0.3175	1.0383
##	[4,]	-0.1	0.2378	-0.1238	-0.0238	0.2293	0.9643
##	[5,]	-0.4	0.2621	-0.3783	0.0217	0.2659	1.0145
##	[6,]	1.0	0.1832	1.0082	0.0082	0.1887	1.0301

##		truth	true_se	estmean_gee_exc	bias_gee_exc	se_gee_exc	re_gee_exc
##	[1,]	-1.2	0.2972	-4.3963	-3.1963	0.4869	1.6382
##	[2,]	-0.4	0.2980	2.1509	2.5509	0.4250	1.4262
##	[3,]	0.4	0.3058	3.1317	2.7317	0.3842	1.2565
##	[4,]	-0.1	0.2378	-2.4585	-2.3585	0.5207	2.1899
##	[5,]	-0.4	0.2621	1.9328	2.3328	0.4828	1.8422
##	[6,]	1.0	0.1832	-0.3843	-1.3843	0.1873	1.0226

Sim2

##		truth	true_se	estmean_gee_ind	bias_gee_ind	se_gee_ind	re_gee_ind
##	[1,]	-1.2	0.3253	-1.2333	-0.0333	0.3269	1.0050
##	[2,]	0.4	0.2604	0.4088	0.0088	0.2555	0.9814
##	[3,]	0.4	0.3194	0.4216	0.0216	0.3259	1.0202
##	[4,]	-0.1	0.2766	-0.1126	-0.0126	0.2688	0.9719
##	[5,]	0.4	0.2944	0.4177	0.0177	0.2936	0.9971
##	[6,]	1.0	0.2170	1.0082	0.0082	0.2254	1.0389

##		truth	true_se	estmean_gee_exc	bias_gee_exc	se_gee_exc	re_gee_exc
##	[1,]	-1.2	0.3253	-4.7498	-3.5498	0.6721	2.0663
##	[2,]	0.4	0.2604	1.8116	1.4116	0.4856	1.8650
##	[3,]	0.4	0.3194	3.4960	3.0960	0.4893	1.5319
##	[4,]	-0.1	0.2766	-2.3612	-2.2612	0.6711	2.4265
##	[5,]	0.4	0.2944	2.5176	2.1176	0.6065	2.0597
##	[6,]	1.0	0.2170	-0.6718	-1.6718	0.3074	1.4168

Sim3

##		truth	true_se	estmean_gee_ind	bias_gee_ind	se_gee_ind	re_gee_ind
##	[1,]	-1.50	0.4394	-1.5638	-0.0638	0.4595	1.0457
##	[2,]	-0.40	0.3852	-0.4245	-0.0245	0.3704	0.9614
##	[3,]	0.40	0.3166	0.3851	-0.0149	0.2950	0.9318
##	[4,]	-0.40	0.3088	-0.4124	-0.0124	0.2906	0.9410
##	[5,]	0.40	0.3995	0.4539	0.0539	0.4169	1.0435
##	[6,]	-0.75	0.3807	-0.7653	-0.0153	0.3762	0.9881
##	[7,]	-0.40	0.2801	-0.4140	-0.0140	0.2813	1.0044
##	[8,]	0.40	0.2910	0.4112	0.0112	0.2887	0.9919
##	[9,]	-0.40	0.3426	-0.4089	-0.0089	0.3322	0.9699
##	[10,]	0.00	0.3330	0.0005	0.0005	0.3313	0.9949
##	[11,]	-0.40	0.2486	-0.4272	-0.0272	0.2435	0.9793
##	[12,]	0.40	0.3165	0.4270	0.0270	0.3083	0.9743
##	[13,]	0.75	0.3284	0.7684	0.0184	0.3235	0.9852
##	[14,]	-0.40	0.3229	-0.3954	0.0046	0.3176	0.9836
##	[15,]	1.50	0.3176	1.5169	0.0169	0.3189	1.0042

##		truth	true_se	estmean_gee_exc	bias_gee_exc	se_gee_exc	re_gee_exc
----	--	-------	---------	-----------------	--------------	------------	------------

##	[1,]	-1.50	0.4394	-5.0473	-3.5473	0.8046	1.8309
##	[2,]	-0.40	0.3852	0.4381	0.8381	0.3311	0.8595
##	[3,]	0.40	0.3166	1.5828	1.1828	0.3338	1.0541
##	[4,]	-0.40	0.3088	0.8646	1.2646	0.3326	1.0769
##	[5,]	0.40	0.3995	2.6236	2.2236	0.8562	2.1429
##	[6,]	-0.75	0.3807	-3.8809	-3.1309	0.4556	1.1967
##	[7,]	-0.40	0.2801	0.6118	1.0118	0.3315	1.1839
##	[8,]	0.40	0.2910	1.9740	1.5740	0.3139	1.0785
##	[9,]	-0.40	0.3426	1.3557	1.7557	0.5304	1.5484
##	[10,]	0.00	0.3330	-1.6660	-1.6660	0.3473	1.0429
##	[11,]	-0.40	0.2486	0.2844	0.6844	0.2976	1.1971
##	[12,]	0.40	0.3165	1.3952	0.9952	0.3345	1.0571
##	[13,]	0.75	0.3284	-0.9469	-1.6969	0.3097	0.9432
##	[14,]	-0.40	0.3229	0.5070	0.9070	0.3461	1.0719
##	[15,]	1.50	0.3176	-0.2062	-1.7062	0.3080	0.9698