

To appear in the *Journal of Statistical Computation and Simulation*
Vol. 00, No. 00, Month 20XX, 1–1

Web-based supporting materials for “A covariance correction that accounts for correlation estimation to improve finite-sample inference with generalized estimating equations: A study on its applicability with structured correlation matrices”

(Received 00 Month 20XX; final version received 00 Month 20XX)

This supporting material presents additional simulation results. The first two tables correspond to Tables 1 and 2 of the manuscript. However, here the true correlation structure is AR-1, rather than exchangeable, with $\alpha_{AR-1} = 0.5$. Because each working correlation structure is equivalent when $n = 2$, these settings are not replicated here.

Tables 3-6 are the results (ESDs, SE estimates, and CPs) from settings in which outcomes are binary and the true structure is either exchangeable or AR-1 with correlation value of 0.2. Tables 7 and 8 present correlation selection results, for which both versions of the CIC utilize the model-based covariance matrix, $\widehat{\Sigma}_I$, that sets $\phi = 1$ because outcomes are binary. Due to lack of stability of the Toeplitz structure, we do not present results for $N < 25$.

References

- [1] Kauermann G, Carroll RJ. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*. 2001;96:1387–1396.
- [2] Mancl LA, DeRouen TA. A covariance estimator for gee with improved small-sample properties. *Biometrics*. 2001;57:126–134.

Table 1. Empirical standard deviations (ESD) (in bold, underneath $\hat{\beta}_1$), empirical mean standard error (SE) estimates (underneath $\hat{\beta}_1$) and corresponding empirical 95% confidence interval coverage probabilities (CP) for both Theoretical and Realistic GEE Analyses. The Kauermann and Carroll [1] correction is utilized, the true structure is AR-1, and outcomes are normal.

Working Correlation	N	n	SE Estimator	Theoretical Analyses		Realistic Analyses	
				$\hat{\beta}_1$	CP	$\hat{\beta}_1$	CP
Exchangeable	10	4	ESD	0.514		0.531	
			SE _T	0.470	0.944	0.469	0.937
			SE _R			0.488	0.946
	50	4	ESD	0.220		0.221	
			SE _T	0.216	0.943	0.216	0.944
			SE _R			0.217	0.947
AR-1	10	4	ESD	0.486		0.499	
			SE _T	0.437	0.939	0.436	0.936
			SE _R			0.456	0.943
	50	4	ESD	0.206		0.207	
			SE _T	0.200	0.948	0.200	0.944
			SE _R			0.202	0.947
Toeplitz	10	4	ESD	—		—	
			SE _T	—	0.939	—	0.890
			SE _R			—	0.948
	25	4	ESD	0.282		0.300	
			SE _T	0.279	0.951	0.279	0.936
			SE _R			0.298	0.949
50	4	ESD	0.202		0.208		
		SE _T	0.199	0.943	0.199	0.937	
		SE _R			0.205	0.947	

N - number of independent subjects; n - number of repeated measurements per subject
 SE_T, theoretical SE estimate obtained from $\hat{\Sigma}$ that assumes correlation parameters are known
 SE_R, realistic SE estimate obtained from $(I_p + \hat{G})\hat{\Sigma}(I_p + \hat{G})^T$
 GEE-generalized estimating equations
 Theoretical Analyses use $\hat{\alpha}(\beta)$ within Equation (1)
 Realistic Analyses use $\hat{\alpha}(\hat{\beta})$ within Equation (1)

Table 2. Empirical standard deviations (ESD) (in bold, underneath $\hat{\beta}_1$), empirical mean standard error (SE) estimates (underneath $\hat{\beta}_1$) and corresponding empirical 95% confidence interval coverage probabilities (CP) for both Theoretical and Realistic GEE Analyses. The Mancl and DeRouen [2] correction is utilized, the true structure is AR-1, and outcomes are normal.

Working Correlation	N	n	SE Estimator	Theoretical Analyses		Realistic Analyses	
				$\hat{\beta}_1$	CP	$\hat{\beta}_1$	CP
Exchangeable	10	4	ESD	0.514		0.531	
			SE _T	0.515	0.963	0.514	0.956
			SE _R			0.537	0.964
	50	4	ESD	0.220		0.221	
			SE _T	0.219	0.950	0.219	0.949
			SE _R			0.221	0.951
AR-1	10	4	ESD	0.486		0.499	
			SE _T	0.481	0.961	0.480	0.936
			SE _R			0.505	0.966
	50	4	ESD	0.206		0.207	
			SE _T	0.204	0.949	0.204	0.948
			SE _R			0.206	0.949
Toeplitz	10	4	ESD	—		—	
			SE _T	—	0.964	—	0.927
			SE _R			—	0.967
	25	4	ESD	0.282		0.300	
			SE _T	0.290	0.957	0.290	0.944
			SE _R			0.311	0.960
50	4	ESD	0.202		0.208		
		SE _T	0.202	0.948	0.202	0.943	
		SE _R			0.209	0.948	

N - number of independent subjects; n - number of repeated measurements per subject
 SE_T, theoretical SE estimate obtained from $\hat{\Sigma}$ that assumes correlation parameters are known
 SE_R, realistic SE estimate obtained from $(I_p + \hat{G})\hat{\Sigma}(I_p + \hat{G})^T$
 GEE-generalized estimating equations
 Theoretical Analyses use $\hat{\alpha}(\beta)$ within Equation (1)
 Realistic Analyses use $\hat{\alpha}(\hat{\beta})$ within Equation (1)

Table 3. Empirical standard deviations (ESD) (in bold, underneath $\hat{\beta}_1$), empirical mean standard error (SE) estimates (underneath $\hat{\beta}_1$) and corresponding empirical 95% confidence interval coverage probabilities (CP) for both Theoretical and Realistic GEE Analyses. The Kauermann and Carroll [1] correction is utilized, the true structure is exchangeable, and outcomes are binary.

Working Correlation	N	n	SE Estimator	Theoretical Analyses		Realistic Analyses	
				$\hat{\beta}_1$	CP	$\hat{\beta}_1$	CP
Any	25	2	ESD	1.037		1.080	
			SE _T	1.042	0.974	1.045	0.967
			SE _R			1.096	0.977
	50	2	ESD	0.715		0.728	
			SE _T	0.708	0.953	0.708	0.949
			SE _R			0.723	0.954
Exchangeable	25	4	ESD	0.702		0.712	
			SE _T	0.689	0.954	0.690	0.955
			SE _R			0.701	0.956
	50	4	ESD	0.498		0.502	
			SE _T	0.478	0.952	0.478	0.951
			SE _R			0.481	0.952
AR-1	25	4	ESD	0.726		0.742	
			SE _T	0.701	0.950	0.701	0.948
			SE _R			0.717	0.954
	50	4	ESD	0.507		0.513	
			SE _T	0.486	0.948	0.486	0.946
			SE _R			0.491	0.948
Toeplitz	25	4	ESD	0.690		0.739	
			SE _T	0.672	0.949	0.673	0.932
			SE _R			0.721	0.955
	50	4	ESD	0.492		0.508	
			SE _T	0.472	0.953	0.472	0.946
			SE _R			0.487	0.952

N - number of independent subjects; n - number of repeated measurements per subject
 SE_T, theoretical SE estimate obtained from $\hat{\Sigma}$ that assumes correlation parameters are known
 SE_R, realistic SE estimate obtained from $(I_p + \hat{G})\hat{\Sigma}(I_p + \hat{G})^T$
 GEE-generalized estimating equations
 Theoretical Analyses use $\hat{\alpha}(\beta)$ within Equation (1)
 Realistic Analyses use $\hat{\alpha}(\tilde{\beta})$ within Equation (1)

Table 4. Empirical standard deviations (ESD) (in bold, underneath $\hat{\beta}_1$), empirical mean standard error (SE) estimates (underneath $\hat{\beta}_1$) and corresponding empirical 95% confidence interval coverage probabilities (CP) for both Theoretical and Realistic GEE Analyses. The Mancl and DeRouen [2] correction is utilized, the true structure is exchangeable, and outcomes are binary.

Working Correlation	N	n	SE Estimator	Theoretical Analyses		Realistic Analyses	
				$\hat{\beta}_1$	CP	$\hat{\beta}_1$	CP
Any	25	2	ESD	1.037		1.080	
			SE _T	1.097	0.979	1.010	0.976
			SE _R			1.157	0.981
	50	2	ESD	0.715		0.728	
			SE _T	0.726	0.957	0.726	0.952
			SE _R			0.742	0.959
Exchangeable	25	4	ESD	0.702		0.712	
			SE _T	0.713	0.963	0.714	0.960
			SE _R			0.726	0.961
	50	4	ESD	0.498		0.502	
			SE _T	0.486	0.955	0.486	0.954
			SE _R			0.489	0.955
AR-1	25	4	ESD	0.726		0.742	
			SE _T	0.726	0.958	0.727	0.954
			SE _R			0.744	0.958
	50	4	ESD	0.507		0.513	
			SE _T	0.495	0.949	0.494	0.946
			SE _R			0.500	0.949
Toeplitz	25	4	ESD	0.690		0.739	
			SE _T	0.696	0.959	0.698	0.940
			SE _R			0.749	0.958
	50	4	ESD	0.492		0.508	
			SE _T	0.480	0.959	0.480	0.949
			SE _R			0.496	0.955

N - number of independent subjects; n - number of repeated measurements per subject
 SE_T, theoretical SE estimate obtained from $\hat{\Sigma}$ that assumes correlation parameters are known
 SE_R, realistic SE estimate obtained from $(I_p + \hat{G})\hat{\Sigma}(I_p + \hat{G})^T$
 GEE-generalized estimating equations
 Theoretical Analyses use $\hat{\alpha}(\beta)$ within Equation (1)
 Realistic Analyses use $\hat{\alpha}(\tilde{\beta})$ within Equation (1)

Table 5. Empirical standard deviations (ESD) (in bold, underneath $\hat{\beta}_1$), empirical mean standard error (SE) estimates (underneath $\hat{\beta}_1$) and corresponding empirical 95% confidence interval coverage probabilities (CP) for both Theoretical and Realistic GEE Analyses. The Kauermann and Carroll [1] correction is utilized, the true structure is AR-1, and outcomes are binary.

Working Correlation	N	n	SE Estimator	Theoretical Analyses		Realistic Analyses	
				$\hat{\beta}_1$	CP	$\hat{\beta}_1$	CP
Exchangeable	25	4	ESD	0.730		0.742	
			SE_T	0.712	0.945	0.712	0.944
			SE_R			0.724	0.947
	50	4	ESD	0.495		0.499	
			SE_T	0.490	0.957	0.490	0.954
			SE_R			0.495	0.957
AR-1	25	4	ESD	0.722		0.738	
			SE_T	0.700	0.949	0.701	0.944
			SE_R			0.715	0.946
	50	4	ESD	0.492		0.496	
			SE_T	0.483	0.949	0.483	0.949
			SE_R			0.488	0.949
Toeplitz	25	4	ESD	0.703		0.752	
			SE_T	0.686	0.944	0.687	0.935
			SE_R			0.734	0.946
	50	4	ESD	0.483		0.498	
			SE_T	0.478	0.948	0.478	0.944
			SE_R			0.493	0.950

N - number of independent subjects; n - number of repeated measurements per subject
 SE_T , theoretical SE estimate obtained from $\hat{\Sigma}$ that assumes correlation parameters are known
 SE_R , realistic SE estimate obtained from $(\mathbf{I}_p + \hat{\mathbf{G}})\hat{\Sigma}(\mathbf{I}_p + \hat{\mathbf{G}})^T$
 GEE-generalized estimating equations
 Theoretical Analyses use $\hat{\alpha}(\beta)$ within Equation (1)
 Realistic Analyses use $\hat{\alpha}(\hat{\beta})$ within Equation (1)

Table 6. Empirical standard deviations (ESD) (in bold, underneath $\hat{\beta}_1$), empirical mean standard error (SE) estimates (underneath $\hat{\beta}_1$) and corresponding empirical 95% confidence interval coverage probabilities (CP) for both Theoretical and Realistic GEE Analyses. The Mancl and DeRouen [2] correction is utilized, the true structure is AR-1, and outcomes are binary.

Working Correlation	N	n	SE Estimator	Theoretical Analyses		Realistic Analyses	
				$\hat{\beta}_1$	CP	$\hat{\beta}_1$	CP
Exchangeable	25	4	ESD	0.730		0.742	
			SE_T	0.737	0.957	0.737	0.954
			SE_R			0.750	0.956
	50	4	ESD	0.495		0.499	
			SE_T	0.499	0.961	0.499	0.958
			SE_R			0.503	0.960
AR-1	25	4	ESD	0.722		0.738	
			SE_T	0.725	0.956	0.725	0.952
			SE_R			0.741	0.956
	50	4	ESD	0.492		0.496	
			SE_T	0.491	0.952	0.491	0.952
			SE_R			0.496	0.953
Toeplitz	25	4	ESD	0.703		0.752	
			SE_T	0.711	0.955	0.712	0.941
			SE_R			0.763	0.957
	50	4	ESD	0.483		0.498	
			SE_T	0.486	0.951	0.486	0.949
			SE_R			0.502	0.951

N - number of independent subjects; n - number of repeated measurements per subject
 SE_T , theoretical SE estimate obtained from $\hat{\Sigma}$ that assumes correlation parameters are known
 SE_R , realistic SE estimate obtained from $(\mathbf{I}_p + \hat{\mathbf{G}})\hat{\Sigma}(\mathbf{I}_p + \hat{\mathbf{G}})^T$
 GEE-generalized estimating equations
 Theoretical Analyses use $\hat{\alpha}(\beta)$ within Equation (1)
 Realistic Analyses use $\hat{\alpha}(\hat{\beta})$ within Equation (1)

Table 7. Empirical frequencies of selecting each working correlation structure out of 1,000 replications from the use of the given correlation selection criterion. The Kauermann and Carroll [1] correction is utilized, and outcomes are binary.

True Correlation	N	Criterion	Selection Frequencies			
			Ind	Exch	AR-1	Toeplitz
Exchangeable	25	TECM _T	43	223	135	599
		TECM _R	194	429	175	202
		CIC _T	43	204	117	636
		CIC _R	229	427	164	180
Exchangeable	50	TECM _T	19	248	106	627
		TECM _R	84	556	160	200
		CIC _T	17	208	87	688
		CIC _R	93	579	149	179
AR-1	25	TECM _T	49	143	189	619
		TECM _R	269	200	347	184
		CIC _T	42	106	177	675
		CIC _R	281	198	358	163
AR-1	50	TECM _T	27	113	248	612
		TECM _R	130	193	476	201
		CIC _T	15	96	237	652
		CIC _R	134	183	495	188

N - number of independent subjects

Ind - Independence; Exch - Exchangeable

TECM - 'trace of the empirical covariance matrix' criterion

CIC - 'correlation information criterion'

T - No penalty: The criterion is based on $\hat{\Sigma}$ that theoretically assumes correlation parameters are known

R - Penalty: The criterion is based on $(\mathbf{I}_p + \hat{\mathbf{G}})\hat{\Sigma}(\mathbf{I}_p + \hat{\mathbf{G}})^T$ that realistically accounts for, or penalizes, covariance inflation due to correlation parameter estimation

Table 8. Empirical frequencies of selecting each working correlation structure out of 1,000 replications from the use of the given correlation selection criterion. The Mancl and DeRouen [2] correction is utilized, and outcomes are binary.

True Correlation	N	Criterion	Selection Frequencies			
			Ind	Exch	AR-1	Toeplitz
Exchangeable	25	TECM _T	34	245	135	586
		TECM _R	196	437	172	195
		CIC _T	35	218	111	636
		CIC _R	205	455	161	179
Exchangeable	50	TECM _T	19	252	105	624
		TECM _R	80	568	157	195
		CIC _T	15	216	79	690
		CIC _R	77	601	147	175
AR-1	25	TECM _T	51	146	208	595
		TECM _R	260	216	356	168
		CIC _T	38	114	196	652
		CIC _R	261	214	378	147
AR-1	50	TECM _T	24	123	249	604
		TECM _R	131	199	479	191
		CIC _T	15	99	248	638
		CIC _R	122	199	496	183

N - number of independent subjects

Ind - Independence; Exch - Exchangeable

TECM - 'trace of the empirical covariance matrix' criterion

CIC - 'correlation information criterion'

T - No penalty: The criterion is based on $\hat{\Sigma}$ that theoretically assumes correlation parameters are known

R - Penalty: The criterion is based on $(\mathbf{I}_p + \hat{\mathbf{G}})\hat{\Sigma}(\mathbf{I}_p + \hat{\mathbf{G}})^T$ that realistically accounts for, or penalizes, covariance inflation due to correlation parameter estimation