

COMPSEL User's Guide

1. Description

COMPSEL is a FORTRAN77 program that computes, both analytically and through simulation, the quality (i.e., the expected average criterion score of the selected and assigned candidates) and the diversity (i.e., the adverse impact ratio, AIR) of **complex selection decisions** that are based on using different predictor composites. For more details on the problem solved by the program we refer to Druart and De Corte (in press). The executable code is offered as is, without any guarantee whatsoever.

At present, the program is limited to 15 criteria, 10 predictors, 5 different applicant groups, 50 different criterion preference subgroups, and no more than 5 different criteria within the criterion preference subgroups. The size of the simulation samples is limited to 50000.

2. Input

Note that all input is in free format: Variables or vectors that have a name commencing with the letters I, J, K, L, M, N get INTEGER values. All other variables, vectors and matrices get FLOATING POINT values. See the example input file.

- # 1: **KEY**. To obtain a KEY, send an e-mail to wilfried.decorte@ugent.be
- # 2: **IVC, NWS, NSIM, NSCP(I), I=1,NWS**
 - IVC = 0: Predictor validities are constant across criteria;
IVC = 1: Predictor validities vary across criteria.
 - NWS: Number of predictor weighing systems for which the corresponding selection quality and diversity will be computed.
 - NSIM: Number of simulation samples to be used when computing selection outcomes through simulation ($1 < \text{NSIM} < 50$).
 - NSCP(I), I=1,NWS: Vector of length NWS where the Ith element specifies the number of score patterns for which details of the analytic computation (e.g., augmented criterion estimates), when using the Ith predictor weighing system, will be printed.
- # 3: **NC, NP, NG, PRG(I), I=1,NG**
 - NC: total number of criteria (i.e., jobs, positions a.s.o.); $\text{NC} < 16$.
 - NP: total number of predictors; $\text{NP} < 11$.

- NG: Number of different applicant subpopulations. At present, the program works for $NG = 2$, where the first subpopulation is the minority population and the second subpopulation corresponds to the majority population.
- PRG(I), $I=1,NG$: Vector of length NG with elements that specify the proportional representation of the applicant subpopulation candidates in the total applicant pool. Note that the sum of the $prg(i)$ values must equal 1; the program does not check this!
- # 4: **NSZ, QU(I), I=1,NC.**
 - NSZ: size of the simulation samples; $500 < NSZ < 50001$, but $NSZ < 20001$ is recommended because the execution time grows quadratically with the value of NSZ.
 - QU(I), $I=1,NC$: Vector of length NC with elements specifying the quota (expressed as a proportion of the total applicant pool) required for the different criteria.
- # 5 (and eventually the following NC-1 lines): **PV(I,J), J=1,NP and I=1 or I=1,NC.**
 - In case that $IVC=0$, $I=1$, and $PV(1,J)$ specifies the validity of the Jth predictor.
 - In case that $IVC=1$, $I=1,NC$, and $PV(I,J)$ specifies the the validity of the Jth predictor in predicting criterion I.
- #6: **NCO.**
 - NCO: number of different application patterns.
- # 7: **PRCO(I), I=1,NCO.**
 - PRCO(I), $I=1,NCO$: Vector of length NCO with elements that specify the proportional presence of the different application patterns. Observe that the sum of the PRCO-elements must equal 1; the program does not check this!
- # 8 and the following NCO-1 lines: **NUP(I), ICO(I,J), J=1,NUP(I).**
 - In total NCO lines. The lth line specifies the number of criteria in application pattern I (cf. $NUP(I)$) and the identity of the $NUP(I)$ criteria that are in this lth application pattern (cf. $ICO(I,J)$, $J=1,NUP(I)$).
- #9 and the following NG-1 lines: **ESP(I,J), I=1,NG, J=1,NP**

- The element in the Ith row and Jth column of $ESP(I,J)$ specifies the effect size of predictor J in applicant subpopulation I. Effect size values must be given relative to the majority group. At present, the program works for $NG = 2$, where the first subpopulation is the minority population and the second subpopulation corresponds to the majority population.
- # 10 and the following NP-1 lines: **Set of NP-1 lines specifying $PC(I,J)$ (with $I = 1, NP-1$ and $J = I+1, NP$) the correlation matrix of the NP predictors.** Note that only the strict upper triangular part of the correlation matrix must be specified! (see example)
- # 11: **DTOL, DREL, DRER**
 - DTOL: relative precision estimation augmentation constants. Recommended value: $DTOL = 0.0001$.
 - DREL: relative precision quadrature computations. Recommended value: $DREL = 0.001$.
 - DRER: relative precision computation of multivariate normal probabilities. Recommended value: $0.001 < DRER < 0.01$.
- #12 and following: **Matrix $SCP(I,J)$, $I = 1, \sum_K NSCP(K)$, $J = 1, NP$**
 - The rows of SCP specify the predictor score score patterns for which the composite scores, estimated criterion scores and the augmented criterion scores are requested.
 - Observe that the total number of rows of SCP must equal $\sum_K NSCP(K)$.
- #13 and following: **Predictor weighing systems**
 - Each set of NC rows specifies a predictor weighing system for which the selection outcomes must be computed.
 - The first row of each set specifies the weights of the predictors in forming the first composite, the second row of each set specifies the weights of the predictors in forming the second composite, a.s.o.
 - Note that in total $NWS*NC$ lines must be specified.

4. Sample Input File

Important: in preparing the input file, use a simple text editor such as Notepad, Wordpad or any other standard ASCII producing editor. DO NOT USE TEXT PROCESSING PROGRAMS SUCH AS MS-WORD or WORDPERFECT. Also, when saving

the input file in Notepad, use the option "All Files" in the "Save as type" box. When saving in Wordpad, use the "Text Document-MS-DOS Format" option in the "Save as type" box, and be **aware that Wordpad has the nasty habit of adding the extension .txt to the file name that you specify**. Thus, with Wordpad, if you specify the name of the input file as "MINPUT", the file will in fact be saved as "MINPUT.TXT"; and this is the name that you have to use in the command to run the present programs. Here is a sample input file (without the first line containing the KEY!!!), for the COMPSEL program.

```

0 5 5 2 2 1 1 1
3 4 2 .12 .88
2000 .25 .10 .15
0.510 0.480 0.220 0.320 0.410
6
.1 .1 .15 .20 .20 .25
1 1
1 2
1 3
2 1 2
2 1 3
3 1 2 3
-0.72 -0.31 -0.06 -0.57 -0.04
0.0 0.0 0.0 0.0 0.0
0.310 0.030 0.370
0.260 0.170
0.310
.0001 .001 .001
-.40 -.25 .50 .20
.30 .40 -.50 -.25
-.40 -.25 .50 .20
.30 .40 -.50 -.25
1. -1. 0. 0.5
.0 .0 .0 .0
.3 -.3 .4 -.6
0.007 0.575 0.681 0.000
0.007 0.575 0.681 0.000
0.007 0.575 0.681 0.000
0.485 0.632 0.270 0.008
0.372 0.709 0.190 0.113
0.448 0.673 0.224 0.034
0.07305263 0.00106614 0.99875894 0.00018167
0.00064214 0.01007441 0.98072933 0.00021820
0.00013543 0.00007634 0.92474010 0.00016108
0.00001663 0.00158681 0.94798284 0.20007568
0.00043279 0.09244496 0.69126273 0.00077156
0.87723732 0.84086466 0.37505394 0.00843838
0.88874075 0.61301113 0.68411363 0.37316190
0.26324599 0.26654975 0.99352162 0.36501105

```

0.15631384 0.65526990 0.80203826 0.16456335 0.47741668

5. Running the Program

Suppose you copied the executable code of the program to the C:\sse1 directory on your machine. In that case, the input file must also be saved in the C:\sse1 directory. Next, to run the program, you have to open an MS-DOS Command window. The way to do this varies from one operating system (i.e., XP, Vista, Windows 7 a.s.o.) to the other, and you should use your local "HELP" button when in doubt about this feature.

If the MS-DOS Command window does not automatically open with the prompt C:\>, then you type in the MS-DOS Command window C:, followed by RETURN or ENTER, and your computer will return the C:\> command prompt. Next, you type cd sse1 after the C:\> command prompt, again followed by RETURN or ENTER, and your computer will respond with the C:\sse1> command prompt. Now, you can execute the program by typing compsel < minput > moutput where "minput" is the name of the input file and "moutput" is the name of the output file. At the end of the execution, the PC will return the command prompt C:\sse1>. You can then inspect the output by editing the output file with either Notepad, Wordpad or any other simple editor program.

6. Sample Output and Description

The above input file corresponds to the file compsel.i at the web page <http://users.ugent.be/~wdecorte/software.html>. The corresponding output is the file compsel.o. Here we comment certain features of the compsel.o output file (Note that the simulation results in the output will vary from one implementation of the program to the other because the random seed is in a nonrepeatable state).

DATE: 09/02/2012; TIME: 15:36:25

THE PRESENT CODE IS FOR DEMONSTRATION PURPOSES ONLY!!

```
+++++++
+ COMPSEL +
+++++++
```

Analytical computation of the selection quality and the adverse impact of COMPLEX selection decisions, using different different predictor composites

Program written by Wilfried De Corte, Ghent University, Belgium

The program uses routines from the Slatec library (see <http://www.netlib.org/slatec>), a couple of algorithms from StatLib (see <http://lib.stat.cmu.edu/apstat/>),

and some (adapted) code from Genz to evaluate multivariate normal probabilities (cf. <http://www.math.wsu.edu/math/faculty/genz/homepage>) At present, the program is limited to 15 criteria, 10 predictors, 5 different applicant groups, 50 different criterion preference subgroups, and no more than 5 different criteria within the criterion preference subgroups. The size of the simulation sample is limited to 50000.

PROBLEM SPECIFICATION

Number of sets of predictor composites: 5
 Number of simulation samples: 5
 Size of simulation samples: 2000
 Number of predictor score patterns per set of composites:

Set 1: 2
 Score pattern 1: -0.40 -0.25 0.50 0.20
 Score pattern 2: 0.30 0.40 -0.50 -0.25
 Set 2: 2
 Score pattern 1: -0.40 -0.25 0.50 0.20
 Score pattern 2: 0.30 0.40 -0.50 -0.25
 Set 3: 1
 Score pattern 1: 1.00 -1.00 0.00 0.50
 Set 4: 1
 Score pattern 1: 0.00 0.00 0.00 0.00
 Set 5: 1
 Score pattern 1: 0.30 -0.30 0.40 -0.60

Total number of different criteria: 3
 Number of predictors: 4
 Number of applicant subpopulations: 2
 Proportional representation of subpopulations: 0.12 0.88
 Number of different criterion preference subgroups: 6
 Group 1; Prop.: 0.10; #Pref: 1; Pref: 1
 Group 2; Prop.: 0.10; #Pref: 1; Pref: 2
 Group 3; Prop.: 0.15; #Pref: 1; Pref: 3
 Group 4; Prop.: 0.20; #Pref: 2; Pref: 1 2
 Group 5; Prop.: 0.20; #Pref: 2; Pref: 1 3
 Group 6; Prop.: 0.25; #Pref: 3; Pref: 1 2 3

Predictor validities are stable over the criteria
 Predictor validities

Criterion	Predictor			
	1	2	3	4
1	0.51	0.48	0.22	0.32
2	0.51	0.48	0.22	0.32
3	0.51	0.48	0.22	0.32

Predictor effect sizes (relative to the majority applicant subpopulation)

Subpopulation	Predictor			
	1	2	3	4

1	-0.72	-0.31	-0.06	-0.57
2	0.00	0.00	0.00	0.00

Predictor correlation matrix

Predictor	Predictors			
	1	2	3	4
1	1.00	0.31	0.03	0.37
2	0.31	1.00	0.26	0.17
3	0.03	0.26	1.00	0.31
4	0.37	0.17	0.31	1.00

Calculation precision parameters

Relative precision computation augmentation constants: 0.0001
 Relative precision of quadratures: 0.0010
 Relative precision of multivariate probabilities: 0.0010

PROGRAM OUTPUT

OUTPUT FOR PREDICTOR COMPOSITE SET 1

Predictor weights in the predictor composites

Composite	Predictors			
	1	2	3	4
1	0.0070	0.5750	0.6810	0.0000
2	0.0070	0.5750	0.6810	0.0000
3	0.0070	0.5750	0.6810	0.0000

Details ANALYTIC computation selection outcomes

Maximal and minimal correlation predictor composites: 1.000 1.000
 Validities predictor composites: 0.429 0.429 0.429
 Analytic computation FAILED

Details SIMULATION computation selection outcomes

	Quality	AIR	Quality per criterion		
			1	2	3
			Simulation 1	0.282	0.797
Simulation 2	0.281	0.851	0.322	0.265	0.226
Simulation 3	0.297	0.833	0.375	0.202	0.232
Simulation 4	0.273	0.824	0.308	0.254	0.229
Simulation 5	0.310	0.879	0.376	0.220	0.259
SIMUL. Aver.	0.289	0.837	0.328	0.280	0.229
SIMUL. Min.	0.273	0.797	0.260	0.202	0.200
SIMUL. Max.	0.310	0.879	0.376	0.458	0.259

OUTPUT FOR PREDICTOR COMPOSITE SET 2

Predictor weights in the predictor composites

Composite	Predictors			
	1	2	3	4

1	0.4850	0.6320	0.2700	0.0080
2	0.3720	0.7090	0.1900	0.1130
3	0.4480	0.6730	0.2240	0.0340

Details ANALYTIC computation selection outcomes

Maximal and minimal correlation predictor composites: 0.999 0.992

Valdities predictor composites: 0.613 0.608 0.612

Cutoff value on highest augmented estimated criteria: -0.031

Augmentation constants: 0.005 -0.038 0.000

Results analytic computation per criterion/job

Criterion	Quality	AIR	Selection rates per group	
			1	2
1	0.460	0.579	0.152	0.263
2	0.428	0.587	0.062	0.105
3	0.442	0.600	0.094	0.158

Global complex selection quality and AIR: 0.448 0.587

Computational details for score vectors

Scorevector 1: -0.40 -0.25 0.50 0.20

	Criterion		
	1	2	3
Raw composite score	-0.215	-0.208	-0.229
Stand. composite score	-0.215	-0.209	-0.229
Estim. criterion score	-0.132	-0.127	-0.140
Augm. crit. est. score	-0.127	-0.165	-0.140

Scorevector 2: 0.30 0.40 -0.50 -0.25

	Criterion		
	1	2	3
Raw composite score	0.261	0.272	0.283
Stand. composite score	0.261	0.272	0.283
Estim. criterion score	0.160	0.165	0.173
Augm. crit. est. score	0.165	0.127	0.173

Details SIMULATION computation selection outcomes

	Quality	AIR	Quality per criterion		
			1	2	3
Simulation 1	0.448	0.569	0.459	0.434	0.439
Simulation 2	0.456	0.638	0.423	0.557	0.442
Simulation 3	0.449	0.577	0.466	0.468	0.409
Simulation 4	0.457	0.544	0.463	0.431	0.465
Simulation 5	0.490	0.629	0.461	0.527	0.512

Comparison ANALYTIC vs SIMULATION computation

	Quality	AIR	Quality per criterion		
			1	2	3
ANAL. Comput.	0.448	0.587	0.460	0.428	0.442
SIMUL. Aver.	0.460	0.591	0.455	0.483	0.453
SIMUL. Min.	0.448	0.544	0.423	0.431	0.409
SIMUL. Max.	0.490	0.638	0.466	0.557	0.512

OUTPUT FOR PREDICTOR COMPOSITE SET 3

Predictor weights in the predictor composites

Composite	Predictors			
	1	2	3	4
1	0.0731	0.0011	0.9988	0.0002
2	0.0006	0.0101	0.9807	0.0002
3	0.0001	0.0001	0.9247	0.0002

Details ANALYTIC computation selection outcomes

Maximal and minimal correlation predictor composites: 1.000 0.997

Validities predictor composites: 0.257 0.225 0.220

Cutoff value on highest augmented estimated criteria: -0.003

Augmentation constants: -0.016 -0.003 0.000

Results analytic computation per criterion/job

Criterion	Quality	AIR	Selection rates per group	
			1	2
1	0.367	0.691	0.179	0.260
2	-0.053	1.211	0.118	0.098
3	-0.130	1.217	0.178	0.146

Global complex selection quality and AIR: 0.134 0.944

Computational details for score vectors

Scorevector 1: 1.00 -1.00 0.00 0.50

	Criterion		
	1	2	3
Raw composite score	0.072	-0.009	0.000
Stand. composite score	0.072	-0.009	0.000
Estim. criterion score	0.018	-0.002	0.000
Augm. crit. est. score	0.002	-0.005	0.000

Details SIMULATION computation selection outcomes

	Quality	AIR	Quality per criterion		
			1	2	3
Simulation 1	0.096	1.009	0.322	-0.021	-0.202
Simulation 2	0.159	0.962	0.388	0.044	-0.145
Simulation 3	0.154	0.962	0.333	-0.006	-0.039
Simulation 4	0.105	0.991	0.360	-0.076	-0.198
Simulation 5	0.186	0.888	0.463	-0.031	-0.132

Comparison ANALYTIC vs SIMULATION computation

	Quality	AIR	Quality per criterion		
			1	2	3
ANAL. Comput.	0.134	0.944	0.367	-0.053	-0.130
SIMUL. Aver.	0.140	0.962	0.374	-0.018	-0.143
SIMUL. Min.	0.096	0.888	0.322	-0.076	-0.202
SIMUL. Max.	0.186	1.009	0.463	0.044	0.000

OUTPUT FOR PREDICTOR COMPOSITE SET 4

Predictor weights in the predictor composites

Composite	Predictors			
-----------	------------	--	--	--

	1	2	3	4
1	0.0000	0.0016	0.9480	0.2001
2	0.0004	0.0924	0.6913	0.0008
3	0.8772	0.8409	0.3751	0.0084

Details ANALYTIC computation selection outcomes

Maximal and minimal correlation predictor composites: 0.978 0.459

Validities predictor composites: 0.266 0.273 0.622

Cutoff value on highest augmented estimated criteria: 0.282

Augmentation constants: 0.279 0.245 0.000

Results analytic computation per criterion/job

Criterion	Quality	AIR	Selection rates per group	
			1	2
1	0.017	0.897	0.227	0.253
2	0.116	1.210	0.118	0.098
3	0.737	0.404	0.065	0.162

Global complex selection quality and AIR: 0.253 0.801

Computational details for score vectors

Scorevector 1: 0.00 0.00 0.00 0.00

	Criterion		
	1	2	3
Raw composite score	0.000	0.000	0.000
Stand. composite score	0.000	0.000	0.000
Estim. criterion score	0.000	0.000	0.000
Augm. crit. est. score	0.279	0.245	0.000

Details SIMULATION computation selection outcomes

	Quality	AIR	Quality per criterion		
			1	2	3
Simulation 1	0.283	0.788	0.018	0.104	0.843
Simulation 2	0.222	0.824	-0.025	0.060	0.742
Simulation 3	0.246	0.851	-0.007	0.079	0.779
Simulation 4	0.281	0.806	0.071	0.083	0.762
Simulation 5	0.233	0.824	-0.003	0.012	0.774

Comparison ANALYTIC vs SIMULATION computation

	Quality	AIR	Quality per criterion		
			1	2	3
ANAL. Comput.	0.253	0.801	0.017	0.116	0.737
SIMUL. Aver.	0.253	0.818	0.011	0.067	0.780
SIMUL. Min.	0.222	0.788	-0.025	0.012	0.742
SIMUL. Max.	0.283	0.851	0.071	0.104	0.843

OUTPUT FOR PREDICTOR COMPOSITE SET 5

Predictor weights in the predictor composites

Composite	Predictors			
	1	2	3	4
1	0.8887	0.6130	0.6841	0.3732
2	0.2632	0.2665	0.9935	0.3650

3 0.1563 0.6553 0.8020 0.1646

Details ANALYTIC computation selection outcomes

Maximal and minimal correlation predictor composites: 0.946 0.877

Valdities predictor composites: 0.603 0.447 0.492

Cutoff value on highest augmented estimated criteria: 0.042

Augmentation constants: 0.004 -0.018 0.000

Results analytic computation per criterion/job

Criterion	Quality	AIR	Selection rates per group	
			1	2
1	0.595	0.399	0.107	0.269
2	0.084	0.868	0.088	0.102
3	0.228	0.938	0.142	0.151

Global complex selection quality and AIR: 0.383 0.646

Computational details for score vectors

Scorevector 1: 0.30 -0.30 0.40 -0.60

	Criterion		
	1	2	3
Raw composite score	0.132	0.177	0.072
Stand. composite score	0.079	0.133	0.057
Estim. criterion score	0.047	0.059	0.028
Augm. crit. est. score	0.051	0.041	0.028

Details SIMULATION computation selection outcomes

	Quality	AIR	Quality per criterion		
			1	2	3
Simulation 1	0.301	0.690	0.518	-0.049	0.173
Simulation 2	0.344	0.544	0.535	0.101	0.186
Simulation 3	0.361	0.655	0.621	0.062	0.128
Simulation 4	0.424	0.646	0.617	0.131	0.299
Simulation 5	0.343	0.544	0.519	0.106	0.210

Comparison ANALYTIC vs SIMULATION computation

	Quality	AIR	Quality per criterion		
			1	2	3
ANAL. Comput.	0.383	0.646	0.595	0.084	0.228
SIMUL. Aver.	0.355	0.616	0.562	0.070	0.199
SIMUL. Min.	0.301	0.544	0.518	-0.049	0.128
SIMUL. Max.	0.424	0.690	0.621	0.131	0.299

CPU Time is 1.123 seconds

7. Acknowledgement

When the user reports results obtained by the present program, due reference should be made to De Corte (2012) and Druart and De Corte (in press).

8. References

De Corte, W. (2012). COMPSEL User's Guide.

Druart, C. & De Corte, W. (in press). Designing Pareto-optimal systems for complex selection decisions. *Organizational Research Methods*.