

BMEM V4.0 Manual

November 6, 2009

1 Background

The software BMEM estimates a mediation model specified in Figure 1.

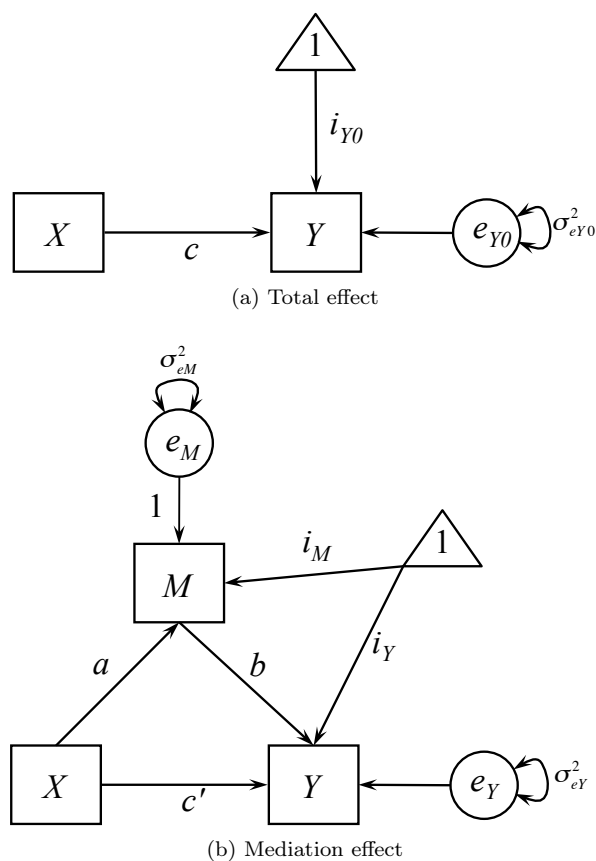


Figure 1: Mediation model

In Figure 1, X , M , and Y are input, mediator, and output variables. The unique feature of BMEM is to test mediation effects with ignorable and non-ignorable missing data. Data can be missed or unobserved on any of the three variables. However, for obvious reasons, there are two requirements for the data:

1. for each case, at least one variable is observed,
2. at least 10 or more cases are complete for all three variables.

Any number of auxiliary variables can be used to explain possible missingness of variables in the mediation model. BMEM implements the Expectation-Maximization (EM) algorithm together with the MLE estimation method for model parameter estimation. To test mediation effects, a bootstrap method is used to

calculate three types of confidence intervals - the percentile interval, the bias-corrected (BC) interval, and the bias corrected and accelerated interval (BCa). Generally speaking, the BCa interval is recommended for use.

2 Availability

BMEM is available as compiled binaries for Windows, OS X, and Linux. The Windows version uses a different sorting algorithm from the one used in OS X and Linux and thus can be slightly slower (bubble sort vs. merge sort). Those binaries can be downloaded from <http://bmem.psychstat.org>. Note that the Mac binary is compiled for Darwin Kernel Version 9.8.0 and the Linux binary is for GNU/Linux Kernel Version 2.6.18-164.e15 and thus those binaries may not work for some versions of OS X and Linux. The current manual focuses on how to use BMEM on PC. Similar procedure applies to OS X and Linux.

3 Files

The following files for the Windows version of BMEM are included in the provided package.

1. BMEM.exe: the executable program
2. BMEM manual.pdf: the current file
3. comp100.txt and comp1000.txt: simulated complete data with sample sizes 100 and 1000 ($a = b = .39$ and $c = 0$)
4. mcar100.txt and mcar1000.txt: simulated MCAR data with sample sizes 100 and 1000 ($a = b = .39$ and $c = 0$)
5. mar100.txt and mar1000.txt: simulated MAR data with sample sizes 100 and 1000 ($a = b = .39$ and $c = 0$)
6. mmar100.txt and mmar1000.txt: simulated MNAR data along with three auxiliary variables with sample sizes 100 and 1000 ($a = b = .39$ and $c = 0$)
7. batch.txt: an example batch file to run BMEM
8. run.bat: a command file to run BMEM in batch mode.

4 How to use

There are two ways to use BMEM – the step by step method and the batch method. For both methods, we suggest putting the data file in the folder where the program file is located.

4.1 Step by step method

After double clicking on the executable file BMEM.exe, a DOS window will pop out and require the input of the following 6 parameters to run the program.

1. The output file name: the name of the file in which one wants to save the results of an analysis.
2. The data file name: the name of a data file. The data file should be a text file with the following sequence of variables, X , M , and Y . If there are auxiliary variables, they should follow the three mediation variables so that the data look like X , M , Y , A_1 , A_2 , ... where A s are auxiliary variables. Variable names are not allowed in the data file.
3. The missing data indicator: the missing data indicator should be a number. An integer such as 88888, 999, or 99999 is recommended.

4. The α level for confidence intervals. It should be a number between 0 and 1. There is no difference to input .05 or .95. BMEM automatically recognizes the confidence level.
5. The random number seed. It should also be a number between 0 and 1. The random number seeds determine the bootstrap samples. The same analysis can be replicated using the identical random number seed.
6. The bootstrap sample size. We suggest a bootstrap sample size of at least 1000.

4.2 Batch method

To use the batch method, one can put the six parameters in the step by step section into a file with each parameter on one line. An example is to create a file called `batch.txt` with the following contents

```
output.txt
input.txt
99999
.95
.9
1000
```

There are two ways to use the batch method. The first way is to use the DOS command. Open the DOS windows through

```
start --> Run... --> cmd
```

In the DOS window, change the directory to where BMEM is located. Then use the command

```
BMEM.exe < batch.txt
```

to run the analysis. The batch file here is saved in the same directory as BMEM. The above batch file conducts the analysis for the data in `input.txt` using the EM algorithm and constructs 95% CIs. All the output is saved in the file `output.txt`.

The second way is more convenient. To help the use of the batch method, a command file `run.bat` is included. To use it, one first change the parameters in the `batch.txt` file and then double click the `run.bat` file. The output is saved in the output file specified in the batch file. To use the method, the data file, `BMEM.exe`, `batch.txt`, and `run.bat` should be in the same folder on the computer.

5 Example output

The following output is from the analysis of the ACTIVE data with one auxiliary variable - V4 (the data may be available on request). There are several important parts of the output.

The first part includes the missing data patterns and the sample size of each pattern. If there are more than 10 patterns, only the first 10 will be shown.

The second part includes the estimated parameters when the mediator is not included in the model (the model in Figure 1a).

The third part includes the estimated parameters when the mediator is included in the model (the model in Figure 1b).

```
-----
| Program name: BMEM.exe (V4.0) |
| See manual.pdf for more information |
-----
```

```
The BMEM program is run on Sat Oct 3 21:37:38 2009
```

```
The output file is: active.edu.txt
The data file is: activeedu.txt
The missing value is: 99999
```

```
-----
| Missing Data Pattern |
-----
```

```
X M Y A1 count
```

```

1 1 1 1 1931
1 0 1 1 194
1 1 0 1 3
1 0 0 1 672
1 1 1 0 1
1 0 1 0 1
The sample size is 2802.
NOTE: 1: observed; 0: missing

```

```

The confidence level is: 0.9
The random number seed is: .9
The bootstrap sample size is: 1000

```

```
-----
```

Model without the mediator									
Parameter	Estimate	S.E.	Percentile		BC		BCa		U
			L	U	L	U	L	U	
iY	36.0067	1.8882	33.0495	38.7784	33.1497	38.8687	33.1497	38.8687	38.8687
c	-0.2287	0.0218	-0.2663	-0.1890	-0.2685	-0.1905	-0.2685	-0.1905	-0.1905
sY2	29.7708	1.3105	27.9302	31.5135	27.9302	31.5135	27.9520	31.5241	31.5241

```
-----
```

Model with the mediator									
Parameter	Estimate	S.E.	Percentile		BC		BCa		U
			L	U	L	U	L	U	
iM	87.2063	4.2853	81.0858	93.7748	81.0920	93.9706	81.0920	93.9706	93.9706
iY	9.9419	1.4345	7.1540	12.6078	7.1749	12.6507	7.1749	12.6507	12.6507
a	-0.7727	0.0502	-0.8614	-0.6899	-0.8641	-0.6909	-0.8622	-0.6909	-0.6909
b	0.2989	0.0120	0.2843	0.3134	0.2843	0.3134	0.2843	0.3134	0.3134
c'	0.0022	0.0177	-0.0316	0.0376	-0.0334	0.0353	-0.0339	0.0351	0.0351
sM2	149.2364	6.4855	141.0218	158.2431	141.2734	158.5305	141.2974	158.5793	158.5793
sY2	16.4390	0.7355	15.4034	17.4300	15.4145	17.4539	15.4471	17.4634	17.4634
a*b	-0.2310	0.0156	-0.2583	-0.2038	-0.2593	-0.2052	-0.2592	-0.2045	-0.2045

NOTE. See the manual for the meaning of each parameter.

The total running time is 77.1329 seconds.

6 Testing

Simulations have been conducted to test BMEM. Most of the functions provided in BMEM are also cross-validated through Mplus for MCAR and MAR data and through an R package for MNAR data that is still under development by us. Here, we present several examples for some naive testing of BMEM.

6.1 Complete data analysis

BMEM certainly can analyze complete data. Simulated data with $a = b = .39; c = 0$ are analyzed and the results are given in Table 1. Note that the results are based on one single replication of simulation. Thus, the parameter estimates may be different from the true values.

Table 1: Results from complete data analysis

			N=100				N=1000			
		TRUE	Est.	S.E.	BCa L	BCa U	Est.	S.E.	BCa L	BCa U
No Mediator	iY		0.050	0.105	-0.151	0.263	0.001	0.033	-0.062	0.067
	c		0.102	0.117	-0.111	0.336	0.160	0.036	0.098	0.228
	sY2		1.114	0.143	0.877	1.453	1.118	0.060	1.032	1.217
With Mediator	iM	0	-0.076	0.091	-0.257	0.101	0.040	0.033	-0.020	0.107
	iY	0	0.076	0.102	-0.129	0.277	-0.014	0.031	-0.074	0.045
	a	0.39	0.354	0.109	0.158	0.577	0.432	0.036	0.365	0.491
	b	0.39	0.347	0.112	0.126	0.555	0.373	0.032	0.318	0.434
	c'	0	-0.021	0.110	-0.221	0.234	-0.001	0.038	-0.076	0.067
	sM2	1	0.825	0.141	0.634	1.216	1.037	0.059	0.949	1.137
	sY2	1	1.015	0.140	0.793	1.322	0.974	0.052	0.902	1.066
	a*b	0.152	0.123	0.059	0.033	0.263	0.161	0.018	0.129	0.196

Note. Only BCa intervals are provided here.

6.2 MCAR data analysis

Data ($N = 100$ and $N = 1000$) are generated under MCAR, MAR, and MNAR for missing data analysis. BMEM can output missing data patterns. Table 2 summarizes the missing data patterns for the MCAR example data. In the output, 1 represents observed data and 0 represents missing data.

Figure 2: Missing data patterns for MCAR example data

Pattern				MCAR	
	X	M	Y	N=100	N=1000
1	1	1	1	50	505
2	0	1	1	14	118
3	1	0	1	8	139
4	0	0	1	3	30
5	1	1	0	15	139
6	0	1	0	4	30
7	1	0	0	6	39

The results from the MCAR data analysis are given in Table 2

Table 2: Results from MAR data analysis

			N=100				N=1000			
		TRUE	Est.	S.E.	BCa L	BCa U	Est.	S.E.	BCa L	BCa U
No Mediator	iY		0.124	0.112	-0.078	0.357	0.008	0.037	-0.056	0.086
	c		-0.094	0.130	-0.355	0.169	0.159	0.042	0.081	0.244
	sY2		0.914	0.132	0.707	1.224	1.115	0.066	1.016	1.246
With Mediator	iM	0	-0.076	0.106	-0.290	0.112	0.075	0.036	0.002	0.143
	iY	0	0.148	0.108	-0.058	0.376	-0.020	0.036	-0.086	0.054
	a	0.39	0.312	0.140	0.053	0.580	0.417	0.041	0.340	0.493
	b	0.39	0.313	0.131	0.035	0.549	0.370	0.041	0.292	0.448
	c'	0	-0.191	0.125	-0.421	0.093	0.005	0.046	-0.083	0.091
	sM2	1	0.948	0.169	0.710	1.387	1.005	0.062	0.908	1.119
	sY2	1	0.821	0.127	0.632	1.143	0.977	0.060	0.893	1.107
	a*b	0.152	0.098	0.065	0.003	0.261	0.155	0.022	0.119	0.201

Note. Only BCa intervals are provided here.

6.3 MAR data analysis

The missing data patterns are given in Table 3.

Figure 3: Missing data patterns for MCAR example data

Pattern	X	M	Y	MAR	
				N=100	N=1000
1	1	1	1	38	468
2	0	1	1	17	155
3	1	0	1	15	120
4	0	0	1	5	39
5	1	1	0	12	112
6	0	1	0	2	31
7	1	0	0	11	75

The results from the MAR data analysis are given in Table 3

Table 3: Results from MAR data analysis

		TRUE	N=100				N=1000			
			Est.	S.E.	BCa L	BCa U	Est.	S.E.	BCa L	BCa U
No Mediator	iY		0.143	0.126	-0.100	0.405	0.010	0.038	-0.061	0.086
	c		0.241	0.162	-0.116	0.533	0.212	0.052	0.113	0.306
	eY2		1.192	0.174	0.918	1.617	1.098	0.066	1.002	1.224
With Mediator	iM	0	-0.107	0.115	-0.385	0.089	-0.041	0.036	-0.110	0.036
	iY	0	0.174	0.130	-0.075	0.434	0.024	0.036	-0.045	0.093
	a	0.39	0.326	0.145	0.040	0.613	0.435	0.043	0.355	0.513
	b	0.39	0.292	0.209	-0.181	0.663	0.349	0.042	0.278	0.436
	c'	0	0.146	0.177	-0.170	0.548	0.060	0.055	-0.050	0.160
	sM2	1	0.760	0.194	0.504	1.379	0.969	0.063	0.858	1.076
	sY2	1	1.127	0.180	0.844	1.551	0.980	0.060	0.891	1.091
	a*b	0.152	0.095	0.086	-0.022	0.366	0.152	0.022	0.113	0.201

Note. Only BCa intervals are provided here.

6.4 MNAR data analysis

The missing data patterns are given in Table 4. A1, A2, and A3 are auxiliary variables.

Figure 4: Missing data patterns for MNAR example data

Pattern	X	M	Y	A1	A2	A3	MNAR	
							N=100	N=1000
1	1	1	1	1	1	1	50	534
2	0	1	1	1	1	1	14	125
3	1	0	1	1	1	1	13	111
4	0	0	1	1	1	1	3	30
5	1	1	0	1	1	1	14	108
6	0	1	0	1	1	1	2	33
7	1	0	0	1	1	1	4	59

The results from the MNAR data analysis are given in Table 4

Table 4: Results from MNAR data analysis

			N=100				N=1000			
		TRUE	Est.	S.E.	BCa L	BCa U	Est.	S.E.	BCa L	BCa U
No Mediator	iY		-0.003	0.130	-0.281	0.237	0.037	0.042	-0.048	0.117
	c		0.041	0.156	-0.252	0.352	0.130	0.044	0.056	0.217
	sY2		1.003	0.162	0.753	1.425	1.187	0.069	1.090	1.316
With Mediator	iM	0	-0.142	0.116	-0.358	0.101	0.029	0.038	-0.045	0.105
	iY	0	0.040	0.122	-0.218	0.265	0.026	0.040	-0.050	0.102
	a	0.39	0.447	0.113	0.201	0.639	0.368	0.043	0.294	0.451
	b	0.39	0.299	0.148	-0.024	0.569	0.368	0.042	0.281	0.440
	c'	0	-0.092	0.182	-0.432	0.265	-0.005	0.043	-0.082	0.086
	sM2	1	0.846	0.120	0.658	1.153	1.012	0.061	0.920	1.129
	sY2	1	0.928	0.150	0.705	1.298	1.050	0.062	0.962	1.176
	a*b	0.152	0.134	0.077	0.012	0.317	0.136	0.021	0.099	0.179

Note. Only BCa intervals are provided here.

6.5 Computing time of BMEM

The EM algorithm and bootstrap method may be time-consuming. Table 5 summarizes computational time in running the examples included in the BMEM package. Note that computational time consumed is related to the sample size and the percentage of missing data.

Table 5: Time consumed in seconds

	Time				Missing percentage		
	Complete	MCAR	MAR	MNAR	MCAR	MAR	MNAR
N=100	0.35	1.80	2.27	10.49	21%	26.7%	19.7%
N=1000	6.33	16.24	27.55	56.60	19.8%	22.6%	19.6%

7 Licence

BMEM is free software. You can redistribute it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 2 of the License, or any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02111-1307 USA

8 Future development

The future development of BMEM will focus on

- Graphical user interface for BMEM
- General mediation models
- Mediator and moderator models
- Longitudinal mediation models

If you are familiar with C++ and are willing to contribute to the development of BMEM, please contact us at xxxx.

9 Citation

It is appreciated for citing the software in the following way if you are willing to use it.

Zhang, Z. & Wang, L. (2009). BMEM (V4.0). Retrievable from <http://bmem.psychstat.org>

10 Acknowledgment

The program is written in C++ to test mediation effects with missing data. Two external libraries are used: Newmat and newran.

1. Davies, R.B. (1994) Writing a matrix package in C++. In OON-SKI'94: The second annual object-oriented numerics conference, pp 207-213. Rogue Wave Software, Corvallis.
2. Eddelbuttel, Dirk (1996) Object-oriented econometrics: matrix programming in C++ using GCC and Newmat. Journal of Applied Econometrics, Vol 11, No 2, pp 199-209.

11 Questions or comments?

Please direct questions or comments to [zhiyongzhang \(a\) nd.edu](mailto:zhiyongzhang (a) nd.edu).