BMEM Manual

Zhiyong Zhang & Lijuan Wang

March 20, 2009

1   Background

The program deals with a mediation model specified as

$$M = iM + aX + e_1$$
$$Y = iY + bM + cX + e_2$$

where X , M , and Y are input, mediator, and output variables. The unique feature of this program is to analyze the mediation effect with missing data. Data can miss on any of the three variables. However, for obvious reasons, we require the data

1. for each case, at least one variable is observed,

2. and at least 10 or more cases are complete for all the three variables.

BMEM implements two types of missing data handling techinques - the pairwise deletion and the Expectation-Maximization (EM) algorithm is used together the MLE estimation method. To assess the mediation effect, either the stratified bootstrap method or the direct bootstrap method can be used to calculate three types of confidence intervals - the percentile interval, the bias-corrected (BC) interval, and the bias corrected and accelerated interval (BCa).

2   Files

The following files are included in the download.

1. BMEM.exe: the exectuable program

2. Manual.pdf: the current file

3. active1.txt: a subset of data from the ACTIVE study

4. active2.txt: a subset of data from the ACTIVE study

5. mar100.txt and mar1000.txt: simulated MAR data with sample sizes 100 and 1000 (a=b=.39 and c=0)

6. mcar100.txt and mcar1000.txt: simulated MCAR data with sample sizes 100 and 1000 (a=b=.39 and c=0)

7. batch.txt: an example batch file to run BMEM

3   How to use

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

3.1 Step by step method

After double clicking on the executable file, a DOS window will pop out and require the following 7 parameters to run the program.

1. The output file name: the name of the file in which one wants to save the analysis results.

2. The data file name: the name of the data file. The data file should be a text file with the following sequence of variables, X, M, and Y. The missing data are should be denoted by 99999, five 9s.

3. The α level for confidence interval. It should be a number between 0 and 1. There is no difference to input .05 or .95. BMEM automatically recognize the level for CI.

4. The random number seed. It should also be between 0 and 1. The random number seeds determine the bootstrap samples. The same analysis can be replicated using the identical random number seed.

5. The bootstrap sample size. We suggest a bootstrap sample size no less than 1000.

6. The missing data handling method. 1 for EM algorithm and 2 for the pairwise deletion method.

7. The bootstrap method. 1 for the stratified bootstrap and 2 for the direct bootstrap.

3.2 Batch method

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

```
output.txt
input.txt
.95
.5
1
1
```

Then, open the DOS windows through `start --> Run... --> cmd`. In the DOS window, change the directory to where BMEM is located. Then using the command `BMEM.ext < 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 using the EM algorithm and stratified bootstrap and constructs the 95% CIs. All the output is saved in the file `output.txt`.

4 Example output

The following output is from the analysis of the attached ACTIVE data (active2.txt). There are several important parts of the output.

The first part is the missing data patterns and sample size of each pattern.

The second part is the results from logistic regression on the test of missingness. For any pair of variable AB, it tests whether A can predict the missing of B. If an abosulte number obtained is larger than 2, one may say the missing data are not MCAR. However, one may not be able to conclude MAR. However, even all numbers are less than 2, one still cannot conlude MCAR.

The third part is the estiamted parameters when the mediator is not considered.

The fourth part is the estimated parameters when the mediator is included in the model.

```
--------------------------------------------------------------
| Program name: BMEM.exe (V3.0)                              |
| See manual.pdf for more information                        |
--------------------------------------------------------------


The BootMed program is run on Fri Mar 20 15:48:46 2009

The output file is: active2.txt
The data file is: active2.txt
The alpha level is: 0.95
The random number seed is: .5


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

Pattern  X  M  Y      Size
1        o  o  o        63
2        o  o  x         3
3        o  x  o        13
4        o  x  x         1
NOTE: o: observed; x: missing




--------------------------------------------------------------
|           Testing Missing Mechanism                        |
--------------------------------------------------------------


        X          M          Y
X             -1.06344  -0.933186
M     N/A              -0.469486
Y     N/A     0.911532
NOTE: N/A means no missing data for B in the pair AB
      Magnitude larger than 2 can be considered as not MCAR


The bootstrap sample size is: 1000
The missing data are handled by EM algorithm.
The boostrap method is stratified bootstrap.


--------------------------------------------------------------
|     Estimated parameters and confidence interval          |
|             EM & STRATIFIED BOOTSTRAP                      |
--------------------------------------------------------------


--------------------------------------------------------------
|                 Model without mediator                     |
--------------------------------------------------------------
```

| | | | Percentile | | BC | | BCa | |
Parameter	Estimate	S.E.	L	U	L	U	L	U
iY	14.7674	1.0239	12.9920	16.7140	13.0000	16.7646	13.0882	16.8196
c	5.2592	1.3067	2.7654	7.6528	2.7654	7.6528	2.6529	7.6199
eY2	47.2724	6.8329	33.1208	59.6471	35.1614	61.6163	36.1410	62.4777

```
--------------------------------------------------------------
|                 Model with mediator                        |
--------------------------------------------------------------
```

| | | | Percentile | | BC | | BCa | |
Parameter	Estimate	S.E.	L	U	L	U	L	U
iM	22.3782	1.8012	19.0931	25.7064	19.0488	25.6339	19.1353	25.7568
iY	4.6776	0.8885	2.8068	6.2040	2.8458	6.2411	2.8458	6.2843
a	12.9178	2.5698	8.0375	18.0439	8.3104	18.2173	8.1839	18.0883
b	0.4509	0.0411	0.3769	0.5326	0.3688	0.5214	0.3683	0.5208
c'	-0.5651	1.1618	-2.7772	1.7479	-2.7083	1.9486	-2.7395	1.7600
eM2	169.8286	24.9932	120.0316	216.8313	125.6172	222.1035	131.2006	236.0085
eY2	12.7480	2.1469	8.3597	16.4475	9.3054	17.8272	9.5787	17.9865
a*b	5.8243	1.1195	3.8455	8.2087	3.9042	8.3292	3.7854	8.1979

```
The total running time is 2.1400 seconds.
```

5   Licence

This is a free program and you can use and distribute it as long as you want. However, we cannot guarantee its performance under any possible circumstances. YOU CAN USE IT FOR FREE BUT AT YOUR OWN RISK. WE ARE NOT RESPONSIBLE FOR ANY LOSS BECAUSE OF THE USE OF BMEM.

6   Citation

Although it is not required, it is appreciated if you can cite the software in the following way if you are willing to use it.

```
Zhang, Z., & Wang, L. (xxxx) Mediation analysis with missing
    data using EM algorithm and bootstrap. xxxx
```

7   Acknowledgment

The program is written in C++ to analyze the meditaion effects with missing data. Two 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.

8   Questions or comments?

Please direct questions or comments to ZhiyongZhang@nd.edu.