```
                                                      ___  ____  ____  ____  ____(R)
                                                     /__    /   ____/   /   ____/
                                                    ___/   /   /___/   /   /___/
                                                      Statistics/Data Analysis

                                                        User: Zara Izadi
                                                     Project: EPI III
> under a causal structure consistent with this hypothesized causal structure.
      name:  <unnamed>
       log:  C:\Users\z_iza\Desktop\PhD\Classes\Epi III\Session 7\mediation.smcl
  log type:  smcl
 opened on:  21 May 2019, 16:02:15

 1 .
 2 . *****Zara Izadi*****
 3 .
 4 . /*
   > Specify a hypothesis regarding a binary exposure, a continuous mediator,
   > and a continuous outcome.  Specify how each variable affects its children
   > i.e., how the exposure influences the mediator) and the distribution of the
   > random or unmeasured determinants of the child variables.
   > */
 5 .
 6 . * Direct effect of x on y, not mediated by m is 1.3
 7 . * Indirect effect of x on y, via m is 1.6 (2*0.8)
 8 . * Total effect of x on y is 1.3+1.6=2.9 and there is no interaction between m and x
 9 .
10 . /*Using the software of your choice, generate a population with 10000 people
   > under a causal structure consistent with this hypothesized causal structure.
   > */
11 .
12 . clear

13 . set seed 4321

14 . set obs 100000
   number of observations (_N) was 0, now 100,000

15 . gen x=runiform()<.5

16 . gen m=2*x+rnormal()*1.5

17 . gen y=1.3*x+0.8*m+rnormal()

18 .
19 . /*Use a conventional Baron-Kenny decomposition to estimate the direct and indirect
   > effects of the exposure on the outcome.
   > */
20 .
21 . reg y x
```

Source	SS	df	MS
Model	212553.188	1	212553.188
Residual	243378.547	99,998	2.43383415
Total	455931.735	99,999	4.55936294

Number of obs	=	100,000
F(1, 99998)	=	87332.65
Prob > F	=	0.0000
R-squared	=	0.4662
Adj R-squared	=	0.4662
Root MSE	=	1.5601

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x	2.915892	.009867	295.52	0.000	2.896553	2.935231
_cons	-.0104899	.0069562	-1.51	0.132	-.0241241	.0031442

```
22 . scalar total_effect=_b[x]

23 . reg y x m

      Source |       SS           df       MS      Number of obs   =   100,000
-------------+----------------------------------   F(2, 99997)     >  99999.00
       Model |  355942.712         2  177971.356   Prob > F        =    0.0000
    Residual |  99989.0232    99,997   .99992023   R-squared       =    0.7807
-------------+----------------------------------   Adj R-squared   =    0.7807
       Total |  455931.735    99,999  4.55936294   Root MSE        =    .99996

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.304514   .0076227   171.14   0.000     1.289574    1.319455
           m |   .8003288   .0021135   378.68   0.000     .7961864    .8044711
       _cons |  -.0025224   .0044588    -0.57   0.572    -.0112616    .0062168
------------------------------------------------------------------------------

24 . scalar direct_effect=_b[x]

25 . dis in red "indirect effect is: " round(total_effect-direct_effect,.001)
indirect effect is: 1.611

26 .
27 . /*Now introduce a confounder of the mediator and outcome (C) into your causal
> model.  Define the new causal models and simulate a new data set.
> Use a conventional decomposition without control for the confounder first and
> then with control for the confounder to derive estimates of the direct and
> indirect effects of exposure on outcome.
> */
28 .
29 . clear

30 . set seed 4321

31 . set obs 100000
number of observations (_N) was 0, now 100,000

32 . gen x=runiform()<.5

33 . gen c=rnormal()

34 . gen m=2*x+c+rnormal()*1.5

35 . gen y=1.3*x+0.8*m+0.5*c+rnormal()

36 .
37 . * Using BK:
38 . reg y x

      Source |       SS           df       MS      Number of obs   =   100,000
-------------+----------------------------------   F(1, 99998)     =  51940.34
       Model |  212715.089         1  212715.089   Prob > F        =    0.0000
    Residual |   409529.15    99,998  4.09537341   R-squared       =    0.3419
-------------+----------------------------------   Adj R-squared   =    0.3418
       Total |  622244.239    99,999  6.22250461   Root MSE        =    2.0237

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   2.917003   .0127992   227.90   0.000     2.891916    2.942089
       _cons |  -.0133455   .0090235    -1.48   0.139    -.0310314    .0043405
------------------------------------------------------------------------------
```

```
39 . scalar total_effect_est=_b[x]

40 . reg y m

      Source |       SS           df       MS      Number of obs   =   100,000
-------------+----------------------------------   F(1, 99998)     >  99999.00
       Model |  486946.11         1   486946.11   Prob > F        =    0.0000
    Residual |  135298.129    99,998  1.35300835   R-squared       =    0.7826
-------------+----------------------------------   Adj R-squared   =    0.7826
       Total |  622244.239    99,999  6.22250461   Root MSE        =    1.1632

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           m |   1.068779   .0017815   599.92   0.000     1.065288    1.072271
       _cons |   .3763356   .0040808    92.22   0.000     .3683373     .384334
------------------------------------------------------------------------------

41 . reg y m x

      Source |       SS           df       MS      Number of obs   =   100,000
-------------+----------------------------------   F(2, 99997)     >  99999.00
       Model |  505988.162         2  252994.081   Prob > F        =    0.0000
    Residual |  116256.077    99,997  1.16259565   R-squared       =    0.8132
-------------+----------------------------------   Adj R-squared   =    0.8132
       Total |  622244.239    99,999  6.22250461   Root MSE        =    1.0782

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           m |   .9504966   .0018925   502.25   0.000     .9467874    .9542059
           x |   1.000141   .0078148   127.98   0.000     .9848244    1.015458
       _cons |  -.0034362   .0048078    -0.71   0.475    -.0128594    .0059871
------------------------------------------------------------------------------

42 . scalar biased_direct_effect_est=_b[x]

43 .
44 . * Checking that right answer is obtained with controlling for C:
45 . reg y m x c

      Source |       SS           df       MS      Number of obs   =   100,000
-------------+----------------------------------   F(3, 99996)     >  99999.00
       Model |  523064.792         3  174354.931   Prob > F        =    0.0000
    Residual |   99179.447    99,996  .991834144   R-squared       =    0.8406
-------------+----------------------------------   Adj R-squared   =    0.8406
       Total |  622244.239    99,999  6.22250461   Root MSE        =    .99591

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           m |   .7978567   .0020997   379.99   0.000     .7937413     .801972
           x |   1.303526   .0075794   171.98   0.000      1.28867    1.318381
           c |   .4976456   .0037926   131.21   0.000     .4902121     .505079
       _cons |  -.0017247   .0044407    -0.39   0.698    -.0104285    .0069791
------------------------------------------------------------------------------
```

```
46 . scalar direct_effect_est=_b[x]

47 . dis direct_effect_est
   1.3035256

48 . dis in red "indirect effect estimated w/o control for c is: " round(total_effect_est-biased_dir
   indirect effect estimated w/o control for c is: 1.917

49 . dis in red "indirect effect estimated w/ control for c is: " round(total_effect_est-direct_effe
   indirect effect estimated w/ control for c is: 1.613

50 . dis in red "total effect is: " round(total_effect_est,.001)
   total effect is: 2.917

51 .
52 . ** Checking it w/ paramed
53 .
54 . paramed y ,avar(x) mvar(m) a0(0) a1(1) m(0) yreg(linear) mreg(linear) cvars(c)

      Source |       SS           df       MS      Number of obs   =   100,000
-------------+----------------------------------   F(4, 99995)     >  99999.00
       Model |  523065.141         4  130766.285   Prob > F        =    0.0000
    Residual |   99179.098    99,995  .991840572   R-squared       =    0.8406
-------------+----------------------------------   Adj R-squared   =    0.8406
       Total |  622244.239    99,999  6.22250461   Root MSE        =    .99591

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   1.305609   .0083532   156.30   0.000     1.289236    1.321981
           m |    .798887   .0027249   293.18   0.000     .7935462    .8042277
       _x_X_m |  -.002074   .0034961    -0.59   0.553    -.0089262    .0047783
           c |   .4976642   .0037928   131.21   0.000     .4902305     .505098
       _cons |  -.0017138   .0044408    -0.39   0.700    -.0104177      .00699
------------------------------------------------------------------------------

      Source |       SS           df       MS      Number of obs   =   100,000
-------------+----------------------------------   F(2, 99997)     =  44740.08
       Model |  201314.322         2  100657.161   Prob > F        =    0.0000
    Residual |  224975.302    99,997  2.24982052   R-squared       =    0.4722
-------------+----------------------------------   Adj R-squared   =    0.4722
       Total |  426289.624    99,999  4.26293887   Root MSE        =    1.4999

------------------------------------------------------------------------------
           m |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   2.007758   .0094867   211.64   0.000     1.989164    2.026351
           c |    1.00074   .0047553   210.45   0.000     .9914194     1.01006
       _cons |  -.0037836   .0066882    -0.57   0.572    -.0168923    .0093251
------------------------------------------------------------------------------

             |   Estimate     Std Err    P>|z|    [95% Conf    Interval]
-------------+-----------------------------------------------------------
         cde |  1.3056085   .00835323    0.000    1.2892362    1.3219809
         nde |  1.3056209   .00836205    0.000    1.2892313    1.3220106
         nie |  1.5998073   .00934822    0.000    1.5814848    1.6181298
         mte |  2.9054283   .00984705    0.000     2.886128    2.9247285

   cde:controlled direct effect, nde:natural direct effect, nie:natural indirect effect, mte:margina
```

```
55 .
56 . /*Create a new version of the mediator that represents a badly measured version
   > of that variable, for example by taking the original variable and adding some
   > random noise to it.  Now use that mediator to evaluate the direct and indirect
   > effects.
   > */
57 . gen m_measured=m+rnormal()*0.7

58 . sum

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
           x |    100,000      .49703    .4999937          0          1
           c |    100,000   -.0021984    .9974831  -4.219059   4.520502
           m |    100,000    .9919322    2.064689  -7.076925   9.336992
           y |    100,000    1.436492    2.494495  -8.168898   11.40392
  m_measured |    100,000    .9904762    2.180661  -7.691277   9.988099

59 .
60 . * Redo above
61 . * Use BK
62 . reg y x

      Source |       SS           df       MS      Number of obs   =   100,000
-------------+----------------------------------   F(1, 99998)     =  51940.34
       Model |  212715.089         1  212715.089   Prob > F        =    0.0000
    Residual |   409529.15    99,998  4.09537341   R-squared       =    0.3419
-------------+----------------------------------   Adj R-squared   =    0.3418
       Total |  622244.239    99,999  6.22250461   Root MSE        =    2.0237

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   2.917003   .0127992   227.90   0.000     2.891916    2.942089
       _cons |  -.0133455   .0090235    -1.48   0.139    -.0310314    .0043405
------------------------------------------------------------------------------

63 . scalar total_effect_werror=_b[x]

64 . reg y m_measured

      Source |       SS           df       MS      Number of obs   =   100,000
-------------+----------------------------------   F(1, 99998)     >  99999.00
       Model |  436296.522         1  436296.522   Prob > F        =    0.0000
    Residual |  185947.717    99,998  1.85951436   R-squared       =    0.7012
-------------+----------------------------------   Adj R-squared   =    0.7012
       Total |  622244.239    99,999  6.22250461   Root MSE        =    1.3636

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  m_measured |   .9578664   .0019775   484.39   0.000     .9539905    .9617422
       _cons |   .4877485   .0047362   102.98   0.000     .4784656    .4970313
------------------------------------------------------------------------------
```

```
65 . reg y m_measured x

      Source |       SS           df       MS      Number of obs   =   100,000
-------------+----------------------------------   F(2, 99997)     >  99999.00
       Model |   467045.52         2  233522.76   Prob > F        =    0.0000
    Residual |  155198.719    99,997  1.55203375   R-squared       =    0.7506
-------------+----------------------------------   Adj R-squared   =    0.7506
       Total |  622244.239    99,999  6.22250461   Root MSE        =    1.2458

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  m_measured |   .8250606   .0020382   404.81   0.000     .8210658    .8290554
           x |   1.251197   .0088892   140.76   0.000     1.233775     1.26862
       _cons |  -.0025931    .005555    -0.47   0.641    -.0134809    .0082946
------------------------------------------------------------------------------

66 . scalar biased_direct_effect_werror=_b[x]

67 . reg y m_measured x c

      Source |       SS           df       MS      Number of obs   =   100,000
-------------+----------------------------------   F(3, 99996)     >  99999.00
       Model |  497213.118         3  165737.706   Prob > F        =    0.0000
    Residual |   125031.12    99,996  1.25036122   R-squared       =    0.7991
-------------+----------------------------------   Adj R-squared   =    0.7991
       Total |  622244.239    99,999  6.22250461   Root MSE        =    1.1182

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  m_measured |   .6540932    .002135   306.37   0.000     .6499086    .6582777
           x |   1.590643   .0082725   192.28   0.000     1.574429    1.606857
           c |   .6426292   .0041372   155.33   0.000     .6345203    .6507381
       _cons |  -.0005562    .004986    -0.11   0.911    -.0103287    .0092164
------------------------------------------------------------------------------

68 . scalar direct_effect_est_werror=_b[x]

69 . dis in red "biased direct effect estimated w/o control for c is: " round(biased_direct_effect_w
   biased direct effect estimated w/o control for c is: 1.251

70 . dis in red "direct effect estimated w control for c is: " round(direct_effect_est_werror, 0.001
   direct effect estimated w control for c is: 1.591

71 . dis in red "indirect effect estimated w/o control for c is: " round(total_effect_werror-biased_
   indirect effect estimated w/o control for c is: 1.666

72 . dis in red "indirect effect estimated w/ control for c is: " round(total_effect_werror-direct_e
   indirect effect estimated w/ control for c is: 1.326

73 . dis in red "total effect is: " round(total_effect_werror,.001)
   total effect is: 2.917
```

```
 74 .
 75 . /*Now try estimating the CDE:
    > */
 76 .
 77 . **(1) make an identifier for your data (in stata, "gen id=_n")*/
 78 . gen id=_n

 79 .
 80 . **(2) make 3 copies of every observation (in stata, use "expand 3"); now you
 81 . **have 2 fake copies of each observation and one real copy.
 82 . expand 3
    (200,000 observations created)

 83 .
 84 . **(3) for the first "fake" copy of each observation, set x to 0 and m to 0
 85 . **and y to .  (missing)
 86 . sort id

 87 . by id: gen copy=_n

 88 . replace m=0 if copy==2
    (100,000 real changes made)

 89 . replace x=0 if copy==2
    (49,703 real changes made)

 90 . replace y=. if copy==2
    (100,000 real changes made, 100,000 to missing)

 91 .
 92 . **(4) for the second "fake" copy of each observation, set x to 1, m to 0 and
 93 . ** y to . (missing)
 94 . replace m=0 if copy==3
    (100,000 real changes made)

 95 . replace x=1 if copy==3
    (50,297 real changes made)

 96 . replace y=. if copy==3
    (100,000 real changes made, 100,000 to missing)

 97 .
 98 . **(5) estimate a regression model predicting the outcome as a function of
 99 . **exposure, mediator, the interaction of the exposure and mediator,
100 . **and the mediator-outcome confounder (C), using only the real observations.
101 .
102 . gen mx=m*x

103 . reg y x c m mx if copy==1
```

Source	SS	df	MS			
				Number of obs	=	**100,000**
				F(4, 99995)	>	**99999.00**
Model	**523065.141**	**4**	**130766.285**	Prob > F	=	**0.0000**
Residual	**99179.098**	**99,995**	**.991840572**	R-squared	=	**0.8406**
				Adj R-squared	=	**0.8406**
Total	**622244.239**	**99,999**	**6.22250461**	Root MSE	=	**.99591**

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x	**1.305609**	**.0083532**	**156.30**	**0.000**	**1.289236**	**1.321981**
c	**.4976642**	**.0037928**	**131.21**	**0.000**	**.4902305**	**.505098**
m	**.798887**	**.0027249**	**293.18**	**0.000**	**.7935462**	**.8042277**
mx	**-.002074**	**.0034961**	**-0.59**	**0.553**	**-.0089262**	**.0047783**
_cons	**-.0017138**	**.0044408**	**-0.39**	**0.700**	**-.0104177**	**.00699**

```
104 .
105 . **(6) for the first fake copy of each observation, use the predict statement
106 . **to predict the counterfactual value of y,    setting x to 0 and m to 0
107 . predict cf_y_x0_m0 if copy==2
    (option xb assumed; fitted values)
    (200,000 missing values generated)

108 .
109 . **(7) for the second fake copy of each observation, use the predict statement
110 . **to predict the counterfactual value of y    setting x to 1 and m to 0
111 . predict cf_y_x1_m0 if copy==3
    (option xb assumed; fitted values)
    (200,000 missing values generated)

112 .
113 . **(8) estimate the controlled direct effect of x on y, setting m to 0
114 . sum cf_y_x0_m0 if copy==2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
cf_y_x0_m0	**100,000**	**-.0028079**	**.4964116**	**-2.101388**	**2.247978**

```
115 . scalar mean_cf_y_x0_m0=r(mean)

116 . sum cf_y_x1_m0 if copy==3
```

Variable	Obs	Mean	Std. Dev.	Min	Max
cf_y_x1_m0	**100,000**	**1.302801**	**.4964116**	**-.7957797**	**3.553587**

```
117 . scalar mean_cf_y_x1_m0=r(mean)

118 .
119 . dis in red "Estimated controlled direct effect of x, setting m to 0, is: " round(mean_cf_y_x1_m
    Estimated controlled direct effect of x, setting m to 0, is: 1.306

120 . ** this is the same as paramed output
121 .
122 . /* Bonus hw if you're having fun.
    >
    > Go back to your original data (before you calculated the CDE*/
123 . clear
```

```
124 . set seed 4321

125 . set obs 100000
    number of observations (_N) was 0, now 100,000

126 . gen x=runiform()<.5

127 . gen c=rnormal()

128 . gen m=2*x+c+rnormal()*1.5

129 . gen y=1.3*x+0.8*m+0.5*c+rnormal()

130 .
131 . **(1) make an identifier for your data (in stata, "gen id=_n")
132 . gen id=_n

133 .
134 . **(2) make 3 copies of every observation (in stata, use "expand 3"); now you
135 . **have 2 fake copies of each observation and one real copy.
136 . expand 3
    (200,000 observations created)

137 .
138 . **(3) for the first "fake" copy of each observation, set x to 0 and m to .
139 . **and y to .
140 . sort id

141 . by id: gen copy=_n

142 . replace m=. if copy==2
    (100,000 real changes made, 100,000 to missing)

143 . replace x=0 if copy==2
    (49,703 real changes made)

144 . replace y=. if copy==2
    (100,000 real changes made, 100,000 to missing)

145 .
146 . **(4) for the second "fake" copy of each observation, set x to 1, m to . and
147 . ** y to .
148 . replace m=. if copy==3
    (100,000 real changes made, 100,000 to missing)

149 . replace x=1 if copy==3
    (50,297 real changes made)

150 . replace y=. if copy==3
    (100,000 real changes made, 100,000 to missing)

151 .
152 . **(5) estimate a regression model predicting the mediator as a function of
```

```
153 . **x and c, using only the real observations
154 . gen mx = m*x
    (200,000 missing values generated)

155 . reg m x c if copy==1

          Source |       SS           df       MS      Number of obs   =   100,000
    -------------+----------------------------------   F(2, 99997)     =  44740.08
           Model |  201314.322         2  100657.161   Prob > F        =    0.0000
        Residual |  224975.302    99,997  2.24982052   R-squared       =    0.4722
    -------------+----------------------------------   Adj R-squared   =    0.4722
           Total |  426289.624    99,999  4.26293887   Root MSE        =    1.4999

    ------------------------------------------------------------------------------
               m |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
    -------------+----------------------------------------------------------------
               x |   2.007758   .0094867   211.64   0.000     1.989164    2.026351
               c |    1.00074   .0047553   210.45   0.000     .9914194     1.01006
           _cons |  -.0037836   .0066882    -0.57   0.572    -.0168923    .0093251
    ------------------------------------------------------------------------------

156 .
157 . **(6) predict cf_m_x0 in the first fake copy
158 . *this gives counterfactual m when exposure is set to 0
159 . predict cf_m_x0 if copy==2
    (option xb assumed; fitted values)
    (200,000 missing values generated)

160 .
161 . **(7) predict cf_m_x1 in the second fake copy
162 . *this gives counterfactual m when exposure is set to 1
163 . predict cf_m_x1 if copy==3
    (option xb assumed; fitted values)
    (200,000 missing values generated)

164 .
165 . **(8) estimate a regression model predicting the outcome as a function of
166 . **exposure, mediator, the interaction of the exposure and mediator,
167 . **and the mediator-outcome confounder (C), usin gonly the real observation
168 . reg y x m mx c if copy==1

          Source |       SS           df       MS      Number of obs   =   100,000
    -------------+----------------------------------   F(4, 99995)     >  99999.00
           Model |  523065.141         4  130766.285   Prob > F        =    0.0000
        Residual |   99179.098    99,995  .991840572   R-squared       =    0.8406
    -------------+----------------------------------   Adj R-squared   =    0.8406
           Total |  622244.239    99,999  6.22250461   Root MSE        =    .99591

    ------------------------------------------------------------------------------
               y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
    -------------+----------------------------------------------------------------
               x |   1.305609   .0083532   156.30   0.000     1.289236    1.321981
               m |    .798887   .0027249   293.18   0.000     .7935462    .8042277
              mx |   -.002074   .0034961    -0.59   0.553    -.0089262    .0047783
               c |   .4976642   .0037928   131.21   0.000     .4902305     .505098
           _cons |  -.0017138   .0044408    -0.39   0.700    -.0104177      .00699
    ------------------------------------------------------------------------------
```

```
. 
. **(9) in the first fake copy, set m to cf_m_x0 and set x to 1
. replace m=cf_m_x0 if copy==2
(100,000 real changes made)

. replace x=1 if copy==2
(100,000 real changes made)

. replace mx=m*x if copy==2
(100,000 real changes made)

. 
. **(10) in the second fake copy, set m to cf_m_x1
. replace m=cf_m_x1  if copy==3
(100,000 real changes made)

. replace mx=m*x if copy==3
(100,000 real changes made)

. 
. **(11) in the first fake copy,  predict y based on the regression model
. **in (8), to estimate cf_y_x1_cf_m_x0
. predict cf_y_x1_cf_m_x0 if copy==2
(option xb assumed; fitted values)
(200,000 missing values generated)

. 
. **(12) in the second fake copy, predict y based on the regression model in (8),
. **to estimate cf_y_x1_cf_m_x1
. predict cf_y_x1_cf_m_x1 if copy==3
(option xb assumed; fitted values)
(200,000 missing values generated)

. 
. **(13) estimate the natural indirect effect of x on y, mediated by m
. * Natural indirect effect is when exposure status is not changing
. * and mediator is changing. In this case exposure is set at 1 and we are
. * comparing outcome under potential mediators at exposed and unexposed.
. * Among exposed what would the outcome be if mediator had the potential value
. * under exposure and what would the outcome be if mediator had the potential value
. * under no exposure. Contract of the potential outcomes.
. sum cf_y_x1_cf_m_x0 if copy==2

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
cf_y_x1_cf~0 |    100,000    1.298033    1.291807  -4.163082   7.155231

. scalar mean_cf_y_x1_cf_m_x0=r(mean)

. sum cf_y_x1_cf_m_x1 if copy==3

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
cf_y_x1_cf~1 |    100,000     2.89784    1.291807  -2.563275   8.755038
```

```
197 . scalar mean_cf_y_x1_cf_m_x1=r(mean)

198 .
199 . dis in red "Estimated natural indirect effect of x on y, mediated by m is: " round(mean_cf_y_x1
    Estimated natural indirect effect of x on y, mediated by m is: 1.6

200 .
201 . ** This is the same as the paramed output
202 . *sort copy
203 . *by copy: sum
204 . log close
          name:  <unnamed>
           log:  C:\Users\z_iza\Desktop\PhD\Classes\Epi III\Session 7\mediation.smcl
      log type:  smcl
     closed on:  21 May 2019, 16:02:17
```