_____

```
      name:  <unnamed>
       log:  /Users/agrogan/Desktop/newstuff/causal-modeling/causal-modeling.s
 > mcl
   log type:  smcl
  opened on:   3 Jul 2020, 12:54:58
```

```
 1 .
 2 . ********************************************
 3 . * causal modeling with GSS data
 4 . * using multiple causal modeling approaches
 5 . ********************************************
 6 .
 7 . * get data
 8 .
 9 . use "/Users/agrogan/Box Sync/DATA WAREHOUSE/General Social Survey Panel Data
   > /GSS_panel2010w123_R6 - stata.dta", clear
   ( )

10 .
11 . * id variable
12 .
13 . generate ID = id_1

14 .
15 . * keep only relevant variables
16 .
17 . keep ID satjob_? educ_? race_? incom16_?

18 .
19 . * describe data
20 .
21 . describe
```

```
   Contains data from /Users/agrogan/Box Sync/DATA WAREHOUSE/General Social Surve
   > y Panel Data/GSS_panel2010w123_R6 - stata.dta
    obs:         2,044
    vars:           13                          12 MAR 2018 16:24
    size:        32,704
```

_____

| | storage | display | value | |
variable name	type	format	label	variable label
educ_1	byte	%8.0g	EDUC_1	**educ_1: HIGHEST YEAR OF SCHOOL COMPLETED**
educ_2	byte	%8.0g	EDUC_2	**educ_2: HIGHEST YEAR OF SCHOOL COMPLETED**
educ_3	byte	%8.0g	EDUC_3	**educ_3: HIGHEST YEAR OF SCHOOL COMPLETED**

STaTa®

incom16_1	byte	%8.0g	INCOM16	**incom16_1: RS FAMILY INCOME WHEN**
				16 YRS OLD
incom16_2	byte	%8.0g	V1318_A	**incom16_2: RS FAMILY INCOME WHEN**
				16 YRS OLD
incom16_3	byte	%8.0g	V1319_A	**incom16_3: RS FAMILY INCOME WHEN**
				16 YRS OLD
race_1	byte	%8.0g	RACE_1	**race_1: RACE OF RESPONDENT**
race_2	byte	%8.0g	RACE_2	**race_2: RACE OF RESPONDENT**
race_3	byte	%8.0g	RACE_3	**race_3: RACE OF RESPONDENT**
satjob_1	byte	%8.0g	SATJOB_1	**satjob_1: JOB OR HOUSEWORK**
satjob_2	byte	%8.0g	SATJOB_2	**satjob_2: JOB OR HOUSEWORK**
satjob_3	byte	%8.0g	SATJOB_3	**satjob_3: JOB OR HOUSEWORK**
ID	float	%9.0g		

Sorted by:
   Note: Dataset has changed since last saved.

```
22 .
23 . * codebook for selected variable(s)
24 .
25 . codebook satjob_3
```

───

satjob_3   **satjob_3: JOB OR HOUSEWORK**
───

```
            type:  numeric (byte)
           label:  SATJOB_3

           range:  [1,4]                      units:  1
   unique values:  4                  missing .:  0/2,044
 unique mv codes:  3                missing .*:  1,086/2,044

      tabulation:  Freq.   Numeric  Label
                     483         1  VERY SATISFIED
                     367         2  MOD. SATISFIED
                      69         3  A LITTLE DISSAT
                      39         4  VERY DISSATISFIED
                       4        .d  DK
                   1,073        .i  IAP
                       9        .n  NA
```

```
26 .
27 . *********************************************
28 . * analyses relying on wide data
29 . *********************************************
30 .
31 . * correlation
32 .
33 . pwcorr satjob_3 educ_3, sig
```

	satjob_3	educ_3
satjob_3	1.0000	
educ_3	-0.0774	1.0000
	0.0166	

```
34 .
35 . * regression with 1 IV
36 .
37 . regress satjob_3 educ_3
```

Source	SS	df	MS
Model	3.53828635	1	3.53828635
Residual	586.493062	955	.61412886
Total	590.031348	956	.617187602

Number of obs	= 957
$F(1, 955)$	= 5.76
Prob > F	= 0.0166
R-squared	= 0.0060
Adj R-squared	= 0.0050
Root MSE	= .78366

satjob_3	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
educ_3	-.0216864	.0090349	-2.40	0.017	-.0394169	-.003956
_cons	1.954439	.1297867	15.06	0.000	1.699739	2.209139

STaTa®

```
38 .
39 . * regression with multiple IV's
40 .
41 . regress satjob_3 educ_3 i.race_3 incom16_3
```

Source	SS	df	MS			
Model	5.81703392	4	1.45425848			
Residual	582.580442	946	.615835563			
Total	588.397476	950	.619365765			

Number of obs	=	951
$F(4, 946)$	=	2.36
Prob > F	=	0.0517
R-squared	=	0.0099
Adj R-squared	=	0.0057
Root MSE	=	.78475

satjob_3	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ_3	-.0215151	.0092674	-2.32	0.020	-.0397021	-.0033281
race_3						
black	.1267666	.0708898	1.79	0.074	-.0123528	.2658861
other	.0677238	.0985112	0.69	0.492	-.1256019	.2610495
incom16_3	.0115275	.0280601	0.41	0.681	-.0435398	.0665947
_cons	1.89556	.144649	13.10	0.000	1.61169	2.17943

```
42 .
43 . * propensity score
44 .
45 . * data wrangling
46 .
47 . generate twelve_years_3 = educ_3 >= 12 // 12 or more years of education

48 .
49 . generate twelve_years_2 = educ_2 >= 12 // 12 or more years of education

50 .
51 . generate twelve_years_1 = educ_1 >= 12 // 12 or more years of education
```

STaTa®

```
52 .
53 . * propensity score analysis
54 .
55 . teffects psmatch (satjob_3) (twelve_years_3 incom16_3 i.race_3)

    Treatment-effects estimation              Number of obs      =       952
    Estimator       : propensity-score matching   Matches: requested =        1
    Outcome model  : matching                              min =        1
    Treatment model: logit                                 max =      296
   ─────────────────────┬──────────────────────────────────────────────────────
   > ──
                        │               AI Robust
        satjob_3 │     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interva
   > l]
                        │
   ─────────────────────┼──────────────────────────────────────────────────────
   > ──
    ATE                 │
    twelve_years_3      │
         (1 vs 0) │  -.0410168   .1083808    -0.38   0.705    -.2534393    .17140
   > 57                 │
                        │
   ─────────────────────┴──────────────────────────────────────────────────────
   > ──

56 .
57 . *******************************************
58 . * analyses relying on long data
59 . *******************************************
60 .
61 . * reshape the data
62 .
63 . reshape long satjob_ educ_ incom16_ race_, i(ID) j(wave)
   (note: j = 1 2 3)

    Data                               wide   ->   long
   ───────────────────────────────────────────────────────────────────
    Number of obs.                     2044   ->    6132
    Number of variables                  16   ->       9
    j variable (3 values)                     ->    wave
    xij variables:
              satjob_1 satjob_2 satjob_3   ->    satjob_
                    educ_1 educ_2 educ_3   ->    educ_
          incom16_1 incom16_2 incom16_3   ->    incom16_
                    race_1 race_2 race_3   ->    race_
   ───────────────────────────────────────────────────────────────────
```

STATA®

```
64 .
65 . * multilevel model
66 .
67 . mixed satjob_ wave educ_ incom16_ i.race_ || ID:

   Performing EM optimization:

   Performing gradient-based optimization:

   Iteration 0:   log likelihood =  -4161.775
   Iteration 1:   log likelihood = -4161.7476
   Iteration 2:   log likelihood = -4161.7476

   Computing standard errors:

   Mixed-effects ML regression               Number of obs      =       3,595
   Group variable: ID                        Number of groups   =       1,661

                                             Obs per group:
                                                           min =           1
                                                           avg =         2.2
                                                           max =           3

                                             Wald chi2(5)       =       42.38
   Log likelihood = -4161.7476               Prob > chi2        =      0.0000
```

satjob_	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
wave	-.018625	.014015	-1.33	0.184	-.0460938	.0088439
educ_	-.018976	.0054133	-3.51	0.000	-.0295859	-.008366
incom16_	-.0350535	.0154559	-2.27	0.023	-.0653465	-.0047606
race_						
black	.1695589	.0451171	3.76	0.000	.0811311	.2579868
other	.035975	.0543135	0.66	0.508	-.0704776	.1424276
_cons	2.049073	.0843019	24.31	0.000	1.883845	2.214302

STaTa®

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
ID: Identity				
var(_cons)	.2305185	.0161162	.2009999	.2643722
var(Residual)	.4174209	.0131143	.3924927	.4439323

LR test vs. linear model: <u>chibar2(01) =</u> **322.95**   Prob >= chibar2 = **0.0000**

```
68 .
69 . * fixed effects regression
70 .
71 . xtreg satjob_ wave educ_ incom16_ i.race_, i(ID) fe
```

Fixed-effects (within) regression   Number of obs   =   **3,595**
Group variable: **ID**   Number of groups   =   **1,661**

R-sq: Obs per group:
    within = **0.0052**   min =   **1**
    between = **0.0148**   avg =   **2.2**
    overall = **0.0122**   max =   **3**

                                               F(**5,1929**)   =   **2.03**
corr(u_i, Xb) = **−0.0714**   Prob > F   =   **0.0711**

satjob_	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
wave	−.0237842	.0152551	−1.56	0.119	−.0537023	.006134
educ_	−.0087664	.0158008	−0.55	0.579	−.0397548	.022222
incom16_	−.047186	.0228265	−2.07	0.039	−.0919531	−.0024189
race_						
black	.3226033	.2025604	1.59	0.111	−.0746572	.7198637
other	.0383663	.104807	0.37	0.714	−.1671806	.2439132
_cons	1.928458	.227991	8.46	0.000	1.481323	2.375593
sigma_u	.6861769					
sigma_e	.64822634					
rho	.52841711	(fraction of variance due to u_i)				

F test that all u_i=0: F(**1660, 1929**) = **2.18**   Prob > F = **0.0000**

```
72 .
73 . * "hybrid" model
74 .
75 .
76 . * close log
77 .
78 . log close _all
         name:  <unnamed>
          log:  /Users/agrogan/Desktop/newstuff/causal-modeling/causal-modeling.s
   > mcl
     log type:  smcl
    closed on:   3 Jul 2020, 12:55:02
      ─────────────────────────────────────────────────────────────────────
```