The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–22

Example using seqlogit

Maarten L. Buis
Department of Social Research Methodology
Vrije Universiteit Amsterdam
Amsterdam, the Netherlands
m.buis@fsw.vu.nl

This document shows some output of an example using the `seqlogit` package. It is meant to clarify some of the tricks used in this example. It is not intended to explain the methodology, this is done in (Buis 2008b) and (Buis 2008a), nor is it intended to give details of the syntax, this is done in the help files of `seqlogit` and `seqlogit postestimation`. If you got this document through installing the ancillary files of the `seqlogit` package from `ssc` you should also have a do-file called seqlogit_example.do and a Stata data file called gss.dta. These two file reproduce the example shown here. The example is based on American data, from the General Social Survey, using the sub-samples of white and African American males. I will assume that children in the US face the educational system as represented in figure 1. When they have not finished high school they can decide to leave or finish high school. If they finish high school they can leave, go to junior college or get a bachelor degree. If they get a bachelor degree they can go to graduate school. Notice that it is assumed that students can not go from junior college to a four year college. With this assumption, there is for each level of education only one way through which it could have been achieved and so knowing the highest achieved level of education is enough to reconstruct the entire educational career of the individual.

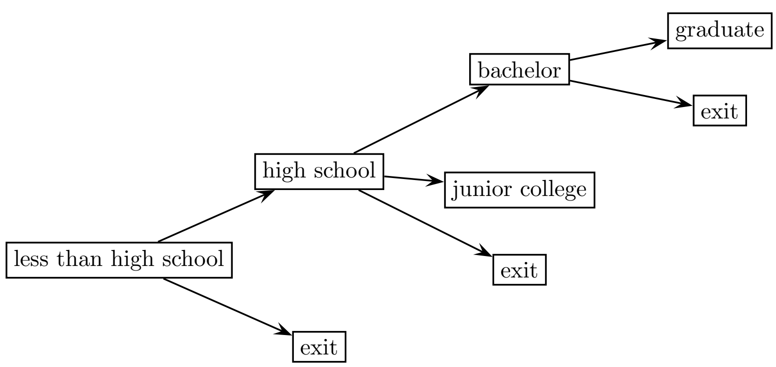

Figure 1: simplified educational system

A short description of the data is shown below.

(*Continued on next page*)


```
.
. desc

Contains data from gss.dta
  obs:        13,421                          General Social Surveys,
                                                1972-2004 [Cumulative File]
 vars:            11                          14 Aug 2007 14:35
 size:       322,104 (96.9% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
sibs            int    %8.0g       SIBS       NUMBER OF BROTHERS AND SISTERS
paeduc          int    %8.0g       PAEDUC     HIGHEST YEAR SCHOOL COMPLETED,
                                                FATHER
degree          int    %8.0g       DEGREE     RS HIGHEST DEGREE
coh             float  %9.0g                  (year of birth-1900)/10
paeducXcoh      float  %9.0g                  interaction term paeduc and coh
black           byte   %8.0g
south           byte   %8.0g                  lived in South when 16 yrs old
country         byte   %8.0g                  lived in the countryside when 16
                                                yrs old
town            byte   %8.0g                  lived in a town when 16 yrs old
suburb          byte   %8.0g                  lived in a suburb when 16 yrs old
city            byte   %8.0g                  lived in a city when 16 yrs old
-------------------------------------------------------------------------------
Sorted by:

. tab degree black, col

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

RS HIGHEST |        black
    DEGREE |         0          1 |     Total
-----------+----------------------+----------
LT HIGH SCHOOL |     1,468        599 |     2,067
           |     13.18      26.52 |     15.42
-----------+----------------------+----------
HIGH SCHOOL |     6,325      1,217 |     7,542
           |     56.77      53.87 |     56.28
-----------+----------------------+----------
JUNIOR COLLEGE |       706        168 |       874
           |      6.34       7.44 |      6.52
-----------+----------------------+----------
  BACHELOR |     1,849        199 |     2,048
           |     16.59       8.81 |     15.28
-----------+----------------------+----------
  GRADUATE |       794         76 |       870
           |      7.13       3.36 |      6.49
-----------+----------------------+----------
     Total |    11,142      2,259 |    13,401
           |    100.00     100.00 |    100.00

.
```

(*Continued on next page*)

A sequential logit is estimated on the sub-sample of white males. Notice how the `tree` option mirrors tree specified in figure 1: The first transition is a choice between less than high school (0) and high school or more (1, 2, 3, and 4), the second transition is a choice between high school (1), junior college (2) and bachelor or graduate (3 and 4), and the final transition is a choice between bachelor and graduate. The main variable of interest is father's education (`paeduc`) and is specified in the `ofinterest()` option. The `over()` option specified that `paeduc` is allowed to change over cohort (`coh`), that is, and interaction effect between `paeduc` and `coh` is added. This way of specifying the explanatory variables is only necessary when one wants to use the post-estimation commands. Some of the post-estimation commands also use the expected value of the highest outcome, which means that if we want to use these command, each educational category needs to be assigned a value. This is done in the `levels()` option. The final line that the results are stored in `white`.

```
. seqlogit degree sibs south country suburb city ///
>    coh if black == 0, or                      ///
>    tree(0:1 2 3 4, 1:2 : 3 4, 3 : 4)          ///
>    ofinterest(paeduc) over(coh)               ///
>    levels(0=9, 1=12, 2=14, 3=16, 4=18)
Transition tree:
Transition 1: 0 : 1 2 3 4
Transition 2: 1 : 2 : 3 4
Transition 3: 3 : 4
Computing starting values for:
Transition 1
Transition 2
Transition 3
Iteration 0:   log likelihood = -9416.7119
Iteration 1:   log likelihood = -9416.7119
```

(*Continued on next page*)

```
                                                  Number of obs   =       8605
                                                  LR chi2(32)     =    2565.13
Log likelihood = -9416.7119                       Prob > chi2     =     0.0000
```

degree	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_1_2_3_4v0						
sibs	.8237616	.0124028	-12.88	0.000	.7998078	.8484329
south	.4984759	.0427117	-8.13	0.000	.4214143	.5896292
country	.8974254	.0841858	-1.15	0.249	.7467044	1.078569
suburb	1.232517	.2061083	1.25	0.211	.8880764	1.710549
city	.8871702	.1240751	-0.86	0.392	.6744695	1.166948
coh	1.350918	.1009637	4.02	0.000	1.166844	1.564031
paeduc	1.210262	.0461014	5.01	0.000	1.123195	1.304077
_paeduc_X_~h	1.000031	.0082445	0.00	0.997	.9840015	1.016321
_2v1						
sibs	.9621188	.0195775	-1.90	0.058	.9245028	1.001265
south	.8950252	.091507	-1.08	0.278	.7325019	1.093608
country	1.042316	.1149873	0.38	0.707	.8396454	1.293908
suburb	1.197037	.1485498	1.45	0.147	.9385888	1.526652
city	.7463013	.1212516	-1.80	0.072	.5427745	1.026145
coh	1.531562	.2023927	3.23	0.001	1.182089	1.984354
paeduc	1.144299	.0674923	2.29	0.022	1.019376	1.284531
_paeduc_X_~h	.9915978	.0109005	-0.77	0.443	.9704616	1.013194
_3_4v1						
sibs	.8447307	.0120464	-11.83	0.000	.8214471	.8686742
south	1.002024	.0631954	0.03	0.974	.8855129	1.133866
country	1.105766	.0793129	1.40	0.161	.9607483	1.272674
suburb	1.35381	.1051141	3.90	0.000	1.1627	1.576332
city	.9937323	.0896516	-0.07	0.944	.832677	1.185939
coh	.828965	.0695354	-2.24	0.025	.7032918	.977095
paeduc	1.142833	.0383614	3.98	0.000	1.070066	1.220549
_paeduc_X_~h	1.020255	.0067988	3.01	0.003	1.007016	1.033667
_4v3						
sibs	.9393345	.0233682	-2.52	0.012	.8946324	.9862703
south	.8919857	.0956887	-1.07	0.287	.7228437	1.100706
country	1.246604	.1584971	1.73	0.083	.9716373	1.599385
suburb	1.292345	.1568546	2.11	0.035	1.018747	1.639421
city	1.648249	.2396358	3.44	0.001	1.239562	2.191682
coh	.7085429	.101275	-2.41	0.016	.5354272	.9376309
paeduc	1.004779	.0515156	0.09	0.926	.9087178	1.110995
_paeduc_X_~h	1.005052	.0102792	0.49	0.622	.9851056	1.025402

```
. estimates store white
```

(*Continued on next page*)

The same is done for the African-American sub-sample.

```
. drop _paeduc_X_coh

.
. seqlogit degree sibs south country suburb city ///
>    coh if black == 1,  or                     ///
>    tree(0:1 2 3 4, 1:2 : 3 4, 3 : 4)          ///
>    ofinterest(paeduc) over(coh)               ///
>    levels(0=9, 1=12, 2=14, 3=16, 4=18)

Transition tree:

Transition 1: 0 : 1 2 3 4
Transition 2: 1 : 2 : 3 4
Transition 3: 3 : 4

Computing starting values for:

Transition 1
Transition 2
Transition 3

Iteration 0:   log likelihood = -1354.7851
Iteration 1:   log likelihood = -1354.7851
```

(*Continued on next page*)

```
                                                  Number of obs   =       1207
                                                  LR chi2(32)     =     277.07
Log likelihood = -1354.7851                       Prob > chi2     =     0.0000
```

degree	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_1_2_3_4v0						
sibs	.8957899	.0247876	-3.98	0.000	.8485011	.9457142
south	.6510122	.133329	-2.10	0.036	.4357734	.9725624
country	.9566217	.1933286	-0.22	0.826	.6437489	1.421556
suburb	1.159284	.4747248	0.36	0.718	.5195466	2.586756
city	1.181978	.2680711	0.74	0.461	.7578058	1.843575
coh	2.271158	.3467367	5.37	0.000	1.683816	3.063373
paeduc	1.221233	.0979997	2.49	0.013	1.043501	1.429238
_paeduc_X_~h	.9742173	.0165611	-1.54	0.124	.942293	1.007223
_2v1						
sibs	1.034587	.0407583	0.86	0.388	.9577084	1.117637
south	1.164525	.2795869	0.63	0.526	.7274225	1.864279
country	.4127671	.1511381	-2.42	0.016	.2013866	.8460178
suburb	1.229247	.4879059	0.52	0.603	.5646571	2.676045
city	.792704	.2096422	-0.88	0.380	.4720612	1.33114
coh	1.424385	.409135	1.23	0.218	.8112066	2.501056
paeduc	1.131095	.1626373	0.86	0.392	.85331	1.499309
_paeduc_X_~h	.9830747	.0255793	-0.66	0.512	.9341971	1.03451
_3_4v1						
sibs	.9292002	.028777	-2.37	0.018	.8744761	.987349
south	1.361952	.2561667	1.64	0.101	.942022	1.969076
country	.636939	.1598802	-1.80	0.072	.3894349	1.041744
suburb	.7789216	.2675193	-0.73	0.467	.3973287	1.526995
city	.7134444	.1474982	-1.63	0.102	.4757534	1.069888
coh	.7273017	.1459607	-1.59	0.113	.4907822	1.077806
paeduc	1.016031	.0940265	0.17	0.864	.8474898	1.218091
_paeduc_X_~h	1.017477	.017949	0.98	0.326	.9828987	1.053272
_4v3						
sibs	1.028093	.0583331	0.49	0.625	.9198897	1.149023
south	1.244753	.4532189	0.60	0.548	.609758	2.541023
country	.74597	.3770551	-0.58	0.562	.2769978	2.008937
suburb	1.960921	1.179725	1.12	0.263	.6030623	6.376143
city	.8949397	.3607191	-0.28	0.783	.4061664	1.971894
coh	.4566354	.1878436	-1.91	0.057	.2038992	1.022642
paeduc	.8478511	.147694	-0.95	0.343	.6026178	1.192881
_paeduc_X_~h	1.036925	.0354648	1.06	0.289	.9696942	1.108818

```
. estimates store black
```

The output shows the log odds ratios of passing each transition. It is also possible to see the odds ratios by specifying the `or` option. The explanatory variable of interest is in this example father's education, and we are interested in seeing how this changed over cohorts. These models not only imply an effect of father's education on passing the different transitions, but also on the highest achieved level of education. As is shown in (Buis 2008b), this effect on the highest achieved level of education is a weighted sum of the log odds ratios. The weights can be obtained with the `predict` command. These weights depend on the values of all explanatory variables so first step is to create a dataset with the appropriate values on the explanatory variables, in this case a person with two siblings, who lived in a town when he was 16 years old (which is the reference category, so `country`, `suburb`, and `city` are all fixed to zero) and not in the south (which is again the reference category, so `south` is fixed at zero) and had an average educated father. Also we only need only one observation per cohort, so the superfluous observations are dropped. This will make the graphs produced by Stata smaller. Next we use `predict` with the `effect` option to predict the effects.

```
. preserve
. sum paeduc, meanonly
. local m = r(mean)
.
. replace sibs= 2
(10892 real changes made)
. replace south = 0
(4716 real changes made)
. replace country = 0
(3326 real changes made)
. replace suburb = 0
(1612 real changes made)
. replace city = 0
(2053 real changes made)
. replace paeduc = `m´
paeduc was int now float
(13421 real changes made)
. replace _paeduc_X_coh = coh*`m´
(13421 real changes made)
.
. sort coh
. by coh: keep if _n == 1
(13355 observations deleted)
.
. estimates restore white
(results white are active now)
. predict effw, effect
```

This procedure is repeated for the African-American sub-sample, and a graph is created.

(*Continued on next page*)

```
. estimates restore black
(results black are active now)
. predict effb, effect
.
. gen byr = coh *10 + 1900
. label variable byr "year of birth"
.
. twoway line effb effw byr,                              ///
>    lpattern(longdash shortdash)                         ///
>    xscale(range(1910 1980)) xlab(1920(20)1980)          ///
>    ytitle("effect of father´s education") name(effect) ///
>    legend(order( 1 "black" 2 "white"))
. restore
```

Notice the difference in trend between the two sub-samples.

Figure 2: Effect of father's education on highest achieved level of education

To see where this difference comes from the effect on the highest achieved level of education is decomposed into the log odds ratios of passing the transitions and the weights assigned to each transition using the `seqlogitdecomp` command. Again we are looking at a person with two siblings, who lived in a town when he was 16 years old, not in the south, and had an average educated father. We are now comparing the cohorts born in 1920, 1930, 1940, 1950, 1960, and 1970. The variable `coh` was coded as (year of birth - 1900)/10, so the values of `coh` we are comparing are 2, 3, ..., 7.

(*Continued on next page*)

```
. estimates restore white
(results white are active now)
. #delimit ;
delimiter now ;
. seqlogitdecomp,
>         overat(coh 2, coh 3, coh 4 , coh 5, coh 6, coh 7)
>         at(sibs 2 south 0 country 0 suburb 0 city 0)
>         subtitle("1920" "1930" "1940" "1950" "1960" "1970")
>         title("white") name(white)
>         eqlabel(
>         `""high school or more" "v. less than high school""´
>         `""junior college" "v. high school or" "bachelor and graduate""´
>         `""bachelor and graduate" "v. high school or" "junior college""´
>         `""graduate" "v. bachelor""´
>         )
>         yscale(range(-.1 .3)) xscale(range(-.1 1.2)) xlabel(0(.5)1)
>         yline(0) xline(0) ;
. #delimit cr
delimiter now cr
```

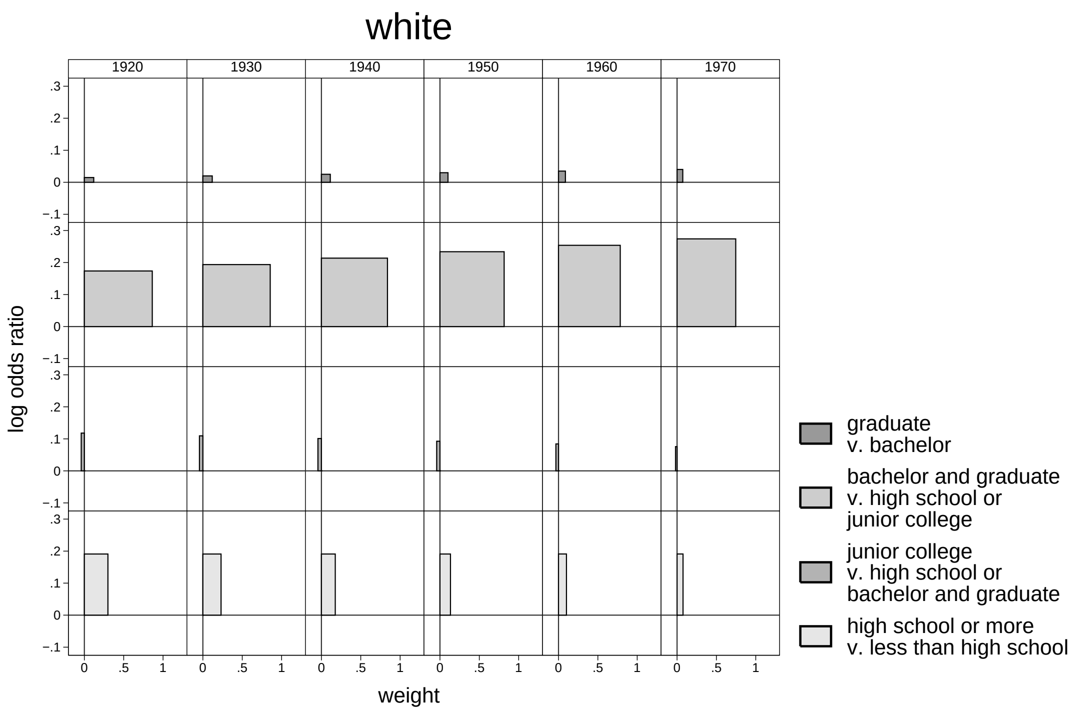

Figure 3: Decomposition of effect on highest achieved level of education into log odds ratios of passing transitions and their weights

This procedure is repeated for the African American sub-sample.

```
. estimates restore black
(results black are active now)

.
. #delimit ;
delimiter now ;
. seqlogitdecomp,
>         overat(coh 2, coh 3, coh 4 , coh 5, coh 6, coh 7)
>         at(sibs 2 south 0 country 0 suburb 0 city 0)
>         subtitle("1920" "1930" "1940" "1950" "1960" "1970")
>         title("black") name(black)
>         eqlabel(
>         `""high school or more" "v. less than high school""´
>         `""junior college" "v. high school or" "bachelor and graduate""´
>         `""bachelor and graduate" "v. high school or" "junior college""´
>         `""graduate" "v. bachelor""´
>         )
>         yscale(range(-.1 .3)) xscale(range(-.1 1.2)) xlabel(0(.5)1)
>         yline(0) xline(0) ;
. #delimit cr
delimiter now cr
```

Notice that for white males the effect on highest achieved level of education is almost entirely the result of the transition between high school and 4 year college (Bachelor), while for African American males the choice whether or not to finish high school was initially the dominant transition. Much of the declining trend for African American males can be explained by decline of this transition.

The weights are the product of three components:

1. the proportion of people at risk of passing the transition, so a transition receives more weight if more people are at risk of passing it.
2. the variance of the dummy indicating whether the transition was passed or not, so a transition receives more weight if close to 50% pass, and less weight if virtually everybody passes or fails that transition.
3. the expected difference in outcome between those that pass and those that fail the transition, so a transition receives more weight if people gain more from passing it.

All these are a function of the probabilities of passing the different transitions, so to see where the differences in weights come from one should first look at the transition probabilities:

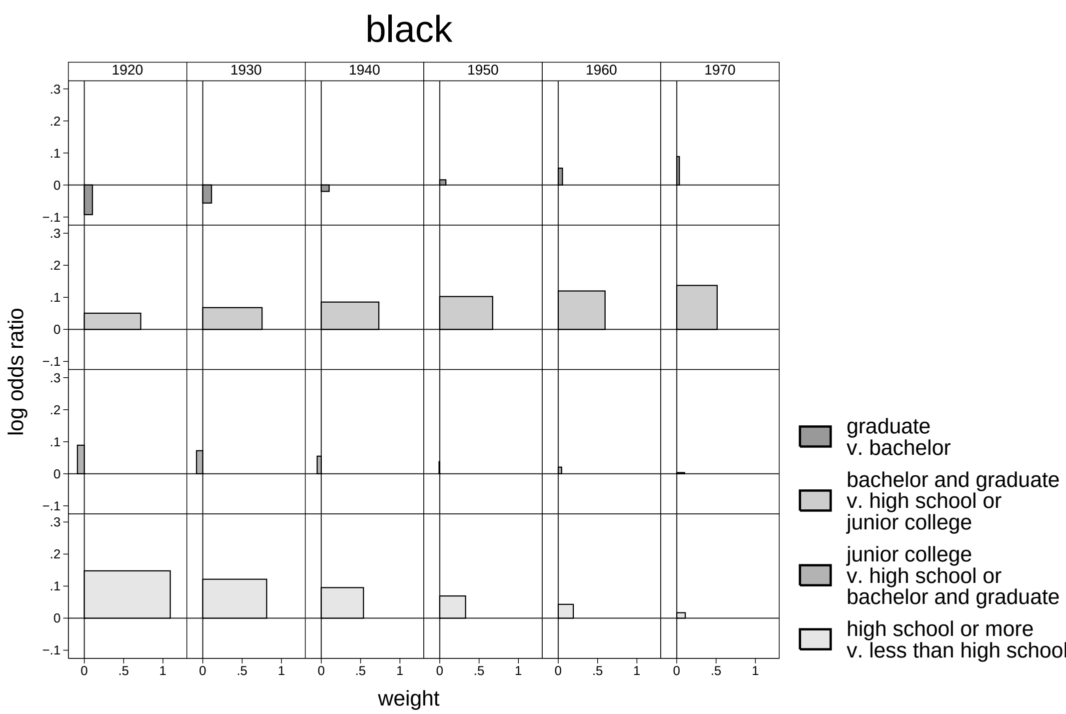

Figure 4: Decomposition of effect on highest achieved level of education into log odds ratios of passing transitions and their weights

```
. preserve
. sum paeduc, meanonly
. local m = r(mean)
.
. replace sibs= 2
(10892 real changes made)
. replace south = 0
(4716 real changes made)
. replace country = 0
(3326 real changes made)
. replace suburb = 0
(1612 real changes made)
. replace city = 0
(2053 real changes made)
. replace paeduc = `m´
paeduc was int now float
(13421 real changes made)
. replace _paeduc_X_coh = coh*`m´
(13421 real changes made)
.
. sort coh
. by coh: keep if _n == 1
(13355 observations deleted)
.
. estimates restore white
(results white are active now)
. predict prw*, trpr
.
. gen byr = coh *10 + 1900
. label variable byr "year of birth"
.
. twoway line prw* byr, ///
>  xscale(range(1910 1980)) xlab(1920(20)1980) ///
>  ytitle("transition probability") name(trw) ///
>  title("white") ///
>  legend(order(  ///
>         1 "high school or more v. less than high school" ///
>         2 "junior college v. high school or bachelor and graduate" ///
>         3 "bachelor and graduate v. high school or junior college" ///
>         4 "graduate v. bachelor") )
.
.
. estimates restore black
(results black are active now)
. predict prb*, trpr
```

(*Continued on next page*)

```
. twoway line prb* byr, ///
>  xscale(range(1910 1980)) xlab(1920(20)1980) ///
>  ytitle("transition probability") name(trb) ///
>  title("black") ///
>  legend(order(  ///
>         1 "high school or more v. less than high school" ///
>         2 "junior college v. high school or bachelor and graduate" ///
>         3 "bachelor and graduate v. high school or junior college" ///
>         4 "graduate v. bachelor") )
.
. grc1leg trw trb, name(trans)
.
```

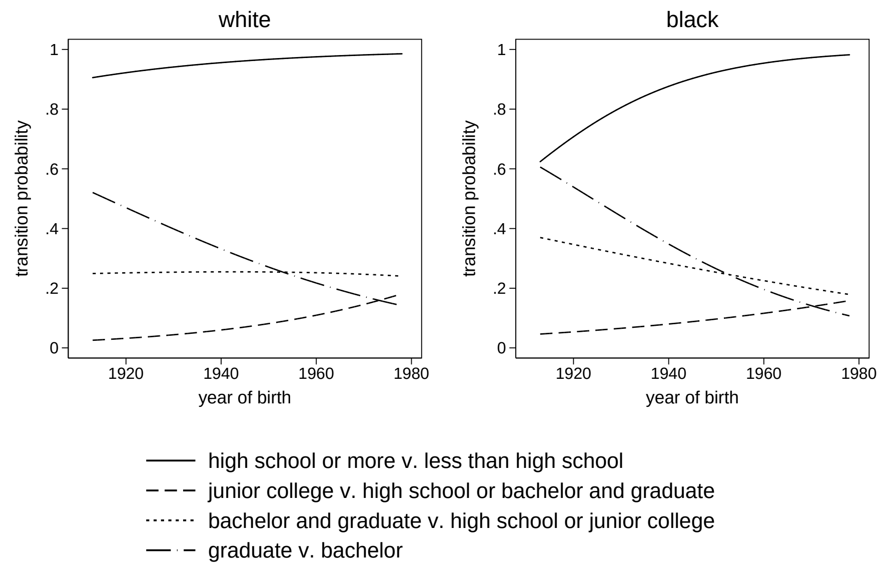

Figure 5: Predicted probabilities of passing transitions

The main difference between the two sub-samples is that initially African Americans where much less likely to finish high school. This difference feeds into the three components of the weights, as can be seen below.

(*Continued on next page*)

```
. preserve

. sum paeduc, meanonly

. local m = r(mean)

.
. replace sibs= 2
(10892 real changes made)

. replace south = 0
(4716 real changes made)

. replace country = 0
(3326 real changes made)

. replace suburb = 0
(1612 real changes made)

. replace city = 0
(2053 real changes made)

. replace paeduc = `m´
paeduc was int now float
(13421 real changes made)

. replace _paeduc_X_coh = coh*`m´
(13421 real changes made)

.
. sort coh

. by coh: keep if _n == 1
(13355 observations deleted)

.
. estimates restore white
(results white are active now)

. predict atriskw*, tratrisk

. predict varw*, trvar

. predict gainw*, trgain

. predict weiw*, trweight

.
. gen byr = coh *10 + 1900

. label variable byr "year of birth"

.
. twoway line atriskw* byr, name(rw) ///
>  xscale(range(1910 1980)) xlab(1920(20)1980) ///
>  legend(order( ///
>        1 "high school or more v." "less than high school" ///
>        2 "junior college v." "high school or" "bachelor and graduate" ///
>        3 "bachelor and graduate v." "high school or" "junior college" ///
>        4 "graduate v." "bachelor") size(vsmall) ) ///
>  ytitle("at risk")

. twoway line varw* byr, name(vw) ///
>   xscale(range(1910 1980)) xlab(1920(20)1980) ///
>   legend(order( ///
>        1 "high school or more v." "less than high school" ///
>        2 "junior college v." "high school or" "bachelor and graduate" ///
>        3 "bachelor and graduate v." "high school or" "junior college" ///
>        4 "graduate v." "bachelor") size(vsmall) ) ///
>   ytitle("variance")
```

(*Continued on next page*)

```
. twoway line gainw* byr, name(gw) ///
>  xscale(range(1910 1980)) xlab(1920(20)1980) ///
>  legend(order( ///
>         1 "high school or more v." "less than high school" ///
>         2 "junior college v." "high school or" "bachelor and graduate" ///
>         3 "bachelor and graduate v." "high school or" "junior college" ///
>         4 "graduate v." "bachelor") size(vsmall) ) ///
>   ytitle("gain")
. twoway line weiw* byr, name(ww) ///
>  xscale(range(1910 1980)) xlab(1920(20)1980) ///
> legend(order( ///
>         1 "high school or more v." "less than high school" ///
>         2 "junior college v." "high school or" "bachelor and graduate" ///
>         3 "bachelor and graduate v." "high school or" "junior college" ///
>         4 "graduate v." "bachelor") size(vsmall) ) ///
>   ytitle("weight")
.
. grc1leg rw vw gw ww, cols(3) holes(4) name(cw) title("white") ring(0) pos(4)
.
. graph export example/txt/ww.eps, replace
(file example/txt/ww.eps written in EPS format)
.
. estimates restore black
(results black are active now)
. predict atriskb*, tratrisk
. predict varb*, trvar
. predict gainb*, trgain
. predict weib*, trweight
.
. twoway line atriskb* byr, name(rb) ///
>  xscale(range(1910 1980)) xlab(1920(20)1980) ///
>  legend(order( ///
>         1 "high school or more v." "less than high school" ///
>         2 "junior college v." "high school or" "bachelor and graduate" ///
>         3 "bachelor and graduate v." "high school or" "junior college" ///
>         4 "graduate v." "bachelor") size(vsmall) ) ///
>   ytitle("at risk")
. twoway line varb* byr, name(vb) ///
>  xscale(range(1910 1980)) xlab(1920(20)1980) ///
>  legend(order( ///
>         1 "high school or more v." "less than high school" ///
>         2 "junior college v." "high school or" "bachelor and graduate" ///
>         3 "bachelor and graduate v." "high school or" "junior college" ///
>         4 "graduate v." "bachelor") size(vsmall) ) ///
>   ytitle("variance")
. twoway line gainb* byr, name(gb) ///
>  xscale(range(1910 1980)) xlab(1920(20)1980) ///
>  legend(order( ///
>         1 "high school or more v." "less than high school" ///
>         2 "junior college v." "high school or" "bachelor and graduate" ///
>         3 "bachelor and graduate v." "high school or" "junior college" ///
>         4 "graduate v." "bachelor") size(vsmall) ) ///
>   ytitle("gain")
```

(*Continued on next page*)

```
. twoway line weib* byr, name(wb) ///
>  xscale(range(1910 1980)) xlab(1920(20)1980) ///
>  legend(order( ///
>         1 "high school or more v." "less than high school" ///
>         2 "junior college v." "high school or" "bachelor and graduate" ///
>         3 "bachelor and graduate v." "high school or" "junior college" ///
>         4 "graduate v." "bachelor") size(vsmall) ) ///
>   ytitle("weight")
.
. grc1leg rb vb gb wb, cols(3) holes(4) name(cb) title("black") ring(0) pos(4)
.
. graph export example/txt/wb.eps, replace
(file example/txt/wb.eps written in EPS format)
.
. restore
```

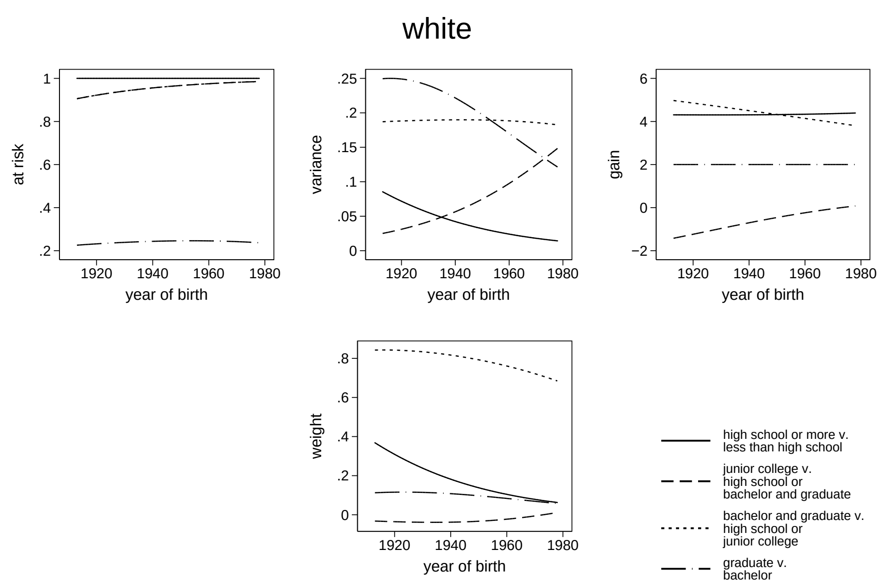

Figure 6: The three components that make up the weights

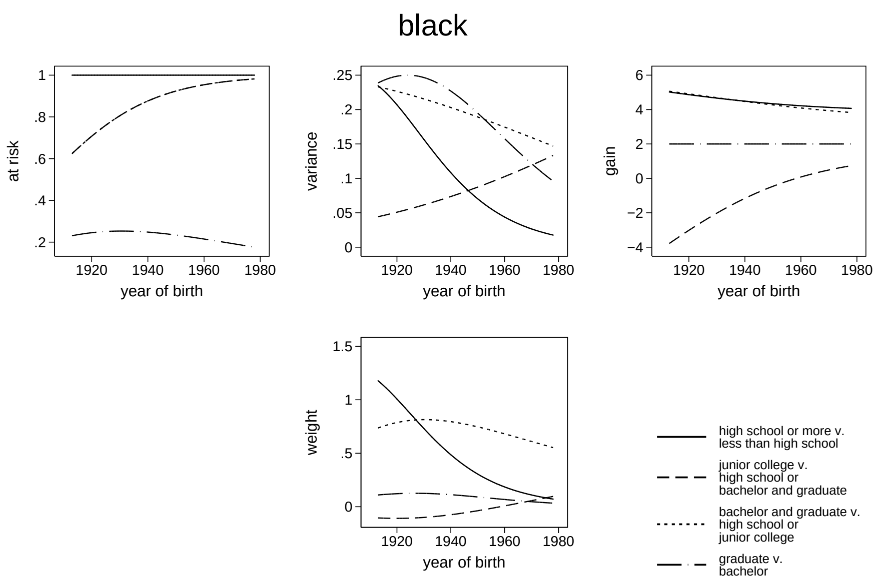

Figure 7: The three components that make up the weights

The `seqlogit` command also contains the `sd()` option to estimate the effect that would occur if one could control for an unobserved normally distributed variable with a standard deviation specified in the `sd()` option, and which is during the first transition uncorrelated with any of the observed variables. This option is intended for performing a sensitivity analysis. To do so, one would re-estimate the model using a number of different reasonable values for the standard deviation, and investigate whether the conclusions change or remain robust. This method is discussed in more detail in (Buis 2008a).

```
. drop _paeduc_X_coh
. seqlogit degree sibs south country suburb city  coh if black == 0, ///
>          or  tree(0:1 2 3 4, 1:2 : 3 4, 3 : 4)  ///
>          ofinterest(paeduc) over(coh) sd(1)
Transition tree:
Transition 1: 0 : 1 2 3 4
Transition 2: 1 : 2 : 3 4
Transition 3: 3 : 4
Computing starting values for:
Transition 1
Transition 2
Transition 3
Iteration 0:   log likelihood = -9576.3256
Iteration 1:   log likelihood =  -9417.072
Iteration 2:   log likelihood = -9416.2433
Iteration 3:   log likelihood = -9416.2433
```

(*Continued on next page*)

```
                                                  Number of obs   =       8605
                                                  LR chi2(32)     =    2566.07
Log likelihood = -9416.2433                       Prob > chi2     =     0.0000
```

degree	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_1_2_3_4v0						
sibs	.8017449	.0136828	-12.95	0.000	.7753708	.8290162
south	.4585072	.0439189	-8.14	0.000	.3800252	.5531971
country	.8939824	.0940679	-1.07	0.287	.7273822	1.098741
suburb	1.289316	.2341737	1.40	0.162	.903148	1.840603
city	.8692876	.1340587	-0.91	0.364	.6425308	1.17607
coh	1.437848	.1228386	4.25	0.000	1.216166	1.699937
paeduc	1.20536	.051681	4.36	0.000	1.108206	1.311031
_paeduc_X_~h	.9952772	.0091567	-0.51	0.607	.9774912	1.013387
_2v1						
sibs	.931144	.0200288	-3.32	0.001	.8927042	.971239
south	.8638969	.0936012	-1.35	0.177	.6986122	1.068286
country	1.051554	.1233927	0.43	0.668	.8355043	1.32347
suburb	1.26571	.1678307	1.78	0.076	.9760373	1.641352
city	.7376786	.1259632	-1.78	0.075	.5278607	1.030897
coh	1.586091	.2202713	3.32	0.001	1.208137	2.082286
paeduc	1.135127	.0698685	2.06	0.039	1.006125	1.280669
_paeduc_X_~h	.9916887	.0114461	-0.72	0.470	.9695067	1.014378
_3_4v1						
sibs	.8180824	.0131051	-12.53	0.000	.7927959	.8441754
south	.9660915	.0701366	-0.48	0.635	.8379583	1.113818
country	1.117765	.0918093	1.36	0.175	.9515595	1.313002
suburb	1.429422	.1292345	3.95	0.000	1.1973	1.706545
city	.9884489	.1030442	-0.11	0.911	.8057826	1.212525
coh	.8516002	.0796486	-1.72	0.086	.7089643	1.022933
paeduc	1.130445	.0427277	3.24	0.001	1.049727	1.217369
_paeduc_X_~h	1.020729	.0076537	2.74	0.006	1.005838	1.03584
_4v3						
sibs	.9028393	.0256562	-3.60	0.000	.8539288	.9545513
south	.8578608	.1060945	-1.24	0.215	.6732028	1.09317
country	1.305278	.1911589	1.82	0.069	.9795887	1.73925
suburb	1.418215	.199573	2.48	0.013	1.076366	1.868634
city	1.769032	.2993412	3.37	0.001	1.269704	2.464729
coh	.6962884	.1145539	-2.20	0.028	.5043695	.961235
paeduc	.9991627	.059441	-0.01	0.989	.8891961	1.122729
_paeduc_X_~h	1.005108	.0118746	0.43	0.666	.9821014	1.028653

```
The standard deviation of the unobserved variable is fixed at 1
```

Due to selection at each transition, the distribution of the unobserved variable will change over the transition. The `uhdesc` command can be used to describe these changes.

(*Continued on next page*)

```
. uhdesc, draws(10) at(south 0 country 0 suburb 0 city 0)
             | p(atrisk)    mean(e)      sd(e)  corr(e,x)
-------------+--------------------------------------------
transition 1 |     1.000     -0.000      1.000     -0.000
transition 2 |     0.958      0.040      0.982     -0.026
transition 3 |     0.218      0.607      0.902     -0.135

.
. uhdesc, overat(coh 1.5, coh 3 , coh 4.5, coh 6, coh 7.5) ///
>         draws(10) overlab(1915 1930 1945 1960 1975) ///
>         at(south 0 country 0 suburb 0 city 0)
             | p(atrisk)    mean(e)      sd(e)  corr(e,x)
-------------+--------------------------------------------
1915         |
 transition1 |     1.000     -0.000      1.000     -0.000
 transition2 |     0.896      0.094      0.963     -0.053
 transition3 |     0.196      0.701      0.896     -0.123
-------------+--------------------------------------------
1930         |
 transition1 |     1.000     -0.000      1.000     -0.000
 transition2 |     0.929      0.065      0.973     -0.039
 transition3 |     0.209      0.660      0.899     -0.129
-------------+--------------------------------------------
1945         |
 transition1 |     1.000     -0.000      1.000     -0.000
 transition2 |     0.953      0.044      0.981     -0.028
 transition3 |     0.217      0.618      0.902     -0.134
-------------+--------------------------------------------
1960         |
 transition1 |     1.000     -0.000      1.000     -0.000
 transition2 |     0.969      0.029      0.987     -0.019
 transition3 |     0.220      0.572      0.903     -0.135
-------------+--------------------------------------------
1975         |
 transition1 |     1.000     -0.000      1.000     -0.000
 transition2 |     0.980      0.019      0.991     -0.013
 transition3 |     0.215      0.519      0.904     -0.130
```

This sensitivity analysis can be extended by investigating what would happen if the unobserved variable was a confounding variable, i.e. if the unobserved variable is correlated with the variable of interest. This hypothetical correlation can be fixed in the `rho()` option, as is illustrated below:

```
. drop _paeduc_X_coh

. seqlogit degree sibs south country suburb city  coh if black == 0, ///
>          or  tree(0:1 2 3 4, 1:2 : 3 4, 3 : 4)  ///
>          ofinterest(paeduc) over(coh) sd(1) rho(.2)

Transition tree:

Transition 1: 0 : 1 2 3 4
Transition 2: 1 : 2 : 3 4
Transition 3: 3 : 4
```

(*Continued on next page*)

```
Computing starting values for:
Transition 1
Transition 2
Transition 3
Iteration 0:   log likelihood = -9576.3256
Iteration 1:   log likelihood =  -9417.072
Iteration 2:   log likelihood = -9416.2433
Iteration 3:   log likelihood = -9416.2433
                                                  Number of obs   =       8605
                                                  Wald chi2(8)    =     719.68
Log likelihood = -9416.2433                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
      degree | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_1_2_3_4v0   |
        sibs |   .8017449   .0136828   -12.95   0.000     .7753708    .8290162
       south |   .4585072   .0439189    -8.14   0.000     .3800252    .5531971
     country |   .8939824   .0940679    -1.07   0.287     .7273822    1.098741
      suburb |   1.289316   .2341737     1.40   0.162      .903148    1.840603
        city |   .8692876   .1340587    -0.91   0.364     .6425308     1.17607
         coh |   1.437848   .1228386     4.25   0.000     1.216166    1.699937
      paeduc |    1.20536    .051681     4.36   0.000     1.108206    1.311031
_paeduc_X_~h |   .9952772   .0091567    -0.51   0.607     .9774912    1.013387
-------------+----------------------------------------------------------------
_2v1         |
        sibs |    .931144   .0200288    -3.32   0.001     .8927042     .971239
       south |   .8638969   .0936012    -1.35   0.177     .6986122    1.068286
     country |   1.051554   .1233927     0.43   0.668     .8355043     1.32347
      suburb |    1.26571   .1678307     1.78   0.076     .9760373    1.641352
        city |   .7376786   .1259632    -1.78   0.075     .5278607    1.030897
         coh |   1.586091   .2202713     3.32   0.001     1.208137    2.082286
      paeduc |   1.135127   .0698685     2.06   0.039     1.006125    1.280669
_paeduc_X_~h |   .9916887   .0114461    -0.72   0.470     .9695067    1.014378
-------------+----------------------------------------------------------------
_3_4v1       |
        sibs |   .8180824   .0131051   -12.53   0.000     .7927959    .8441754
       south |   .9660915   .0701366    -0.48   0.635     .8379583    1.113818
     country |   1.117765   .0918093     1.36   0.175     .9515595    1.313002
      suburb |   1.429422   .1292345     3.95   0.000       1.1973    1.706545
        city |   .9884489   .1030442    -0.11   0.911     .8057826    1.212525
         coh |   .8516002   .0796486    -1.72   0.086     .7089643    1.022933
      paeduc |   1.130445   .0427277     3.24   0.001     1.049727    1.217369
_paeduc_X_~h |   1.020729   .0076537     2.74   0.006     1.005838     1.03584
-------------+----------------------------------------------------------------
_4v3         |
        sibs |   .9028393   .0256562    -3.60   0.000     .8539288    .9545513
       south |   .8578608   .1060945    -1.24   0.215     .6732028     1.09317
     country |   1.305278   .1911589     1.82   0.069     .9795887     1.73925
      suburb |   1.418215    .199573     2.48   0.013     1.076366    1.868634
        city |   1.769032   .2993412     3.37   0.001     1.269704    2.464729
         coh |   .6962884   .1145539    -2.20   0.028     .5043695     .961235
      paeduc |   .9991627    .059441    -0.01   0.989     .8891961    1.122729
_paeduc_X_~h |   1.005108   .0118746     0.43   0.666     .9821014    1.028653
------------------------------------------------------------------------------
The standard deviation of the unobserved variable is fixed at 1
The initial correlation between the unobserved variable and paeduc is fixed at .2
```

```
. uhdesc, draws(10) at(south 0 country 0 suburb 0 city 0)
             | p(atrisk)    mean(e)      sd(e)  corr(e,x)
-------------+-------------------------------------------
transition 1 |     1.000     -0.000      1.000      0.200
transition 2 |     0.958      0.040      0.982      0.168
transition 3 |     0.218      0.607      0.902      0.037
```

References

Buis, M. L. 2008a. The Consequences of Unobserved Heterogeneity in a Sequential Logit Model. http://home.fsw.vu.nl/m.buis/wp/unobserved_het.pdf.

———. 2008b. Not all transitions are equal: The relationship between inequality of educational opportunities and inequality of educational outcomes. http://home.fsw.vu.nl/m.buis/wp/distmare.html.