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)
Example using seqlogit

\[ \text{. desc} \]
Contains data from gss.dta
\begin{tabular}{ll}
obs: & 13,421 \\
vars: & 11 \\
size: & 322,104 (96.9\% of memory free)
\end{tabular}
\[ \text{General Social Surveys, 1972-2004 [Cumulative File]} \]
14 Aug 2007 14:35

\begin{tabular}{llll}
variable name & storage & display type & format & label \\
\hline
sibs & int & \%8.0g & \\
paeduc & int & \%8.0g & \\
degree & int & \%8.0g & \\
coh & float & \%9.0g & \\
paeducXcoh & float & \%9.0g & \\
black & byte & \%8.0g & \\
south & byte & \%8.0g & \\
country & byte & \%8.0g & \\
town & byte & \%8.0g & \\
suburb & byte & \%8.0g & \\
city & byte & \%8.0g & \\
\end{tabular}

variable label
\begin{tabular}{ll}
NUMBER OF BROTHERS AND SISTERS & SIBS \\
HIGHEST YEAR SCHOOL COMPLETED, FATHER & PAEDUC \\
RS HIGHEST DEGREE & DEGREE \\
(\text{year of birth-1900)/10} & \\
interaction term paeduc and coh & \\
lived in South when 16 yrs old & \\
lived in the countryside when 16 yrs old & \\
lived in a town when 16 yrs old & \\
lived in a suburb when 16 yrs old & \\
lived in a city when 16 yrs old & \\
\end{tabular}

Sorted by:
\[ \text{. tab degree black, col} \]

\[ \text{Key} \]
\begin{tabular}{llll}
\hline
\text{frequency} & black & 0 & 1 & Total \\
\hline
\text{RS HIGHEST DEGREE} & & & & \\
LT HIGH SCHOOL & 1,468 & 599 & 2,067 & 15.42 \\
& 13.18 & 26.52 & 15.42 & \\
HIGH SCHOOL & 6,325 & 1,217 & 7,542 & 56.28 \\
& 56.77 & 53.87 & 56.28 & \\
JUNIOR COLLEGE & 706 & 168 & 874 & 6.52 \\
& 6.34 & 7.44 & 6.52 & \\
BACHELOR & 1,849 & 199 & 2,048 & 15.28 \\
& 16.59 & 8.81 & 15.28 & \\
GRADUATE & 794 & 76 & 870 & 6.49 \\
& 7.13 & 3.36 & 6.49 & \\
\hline
Total & 11,142 & 2,259 & 13,401 & \\
& 100.00 & 100.00 & 100.00 & \\
\end{tabular}

(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)
Example using seqlogit

Example using seqlogit

|       | 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 | .8807064 1.710549 |
| city  | .8871702   | .1240751  | -0.86  | 0.392 | .6744695 1.16948 |
| coh   | 1.350918   | .1009637  | 4.02   | 0.000 | 1.166844 1.564031 |
| paeduc | 1.210262   | .0461914  | 5.01   | 0.000 | 1.123195 1.304077 |
| _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 |
| _3_4v1 |            |           |      |      |                      |
| sibs  | .8447307   | .0120464  | -11.83 | 0.000 | .8214471  .8686472 |
| south | 1.002024   | .0631954  | 0.03   | 0.974 | .8855129 1.133866 |
| country | 1.105766   | .0793129  | 1.40   | 0.161 | .9607483 1.272674 |
| suburb | 1.35581    | .1051141  | 3.90   | 0.000 | 1.1627   1.576332 |
| city  | .9937323   | .0896516  | -0.07  | 0.944 | .832677 1.185939 |
| coh   | .829865    | .0695354  | -2.24  | 0.025 | .7032918 .977096 |
| paeduc | 1.142833   | .0383614  | 3.98   | 0.000 | 1.070066 1.220549 |
| _4v3 |            |           |      |      |                      |
| sibs  | .9393345   | .0236882  | -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  .9378309 |
| 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
The same is done for the African-American sub-sample.

```
.drop _paeduc_X_coh

.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:
Transitions 1: 0 : 1 2 3 4
Transitions 2: 1 : 2 : 3 4
Transitions 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)
Example using `seqlogit`

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)](https://doi.org/10.1093/oxfordhb/9780199204697.013.13), 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_k_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)
Example using `seqlogit`

```
. 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.

![Graph showing the effect of father's education on highest achieved level of education](image)

**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 \((\text{year of birth} - 1900)/10\), 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
Example using seqlogit

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) ;
```

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.
Example using seqlogit

. 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
(13356 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)) xlabel(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)
The main difference between the two sub-samples is that initially African Americans were much less likely to finish high school. This difference feeds into the three components of the weights, as can be seen below.
Example using seqlogit

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
(13356 observations deleted)

estimates restore white
(results white are active now)
predict atriskw*, tratrisk
predict varw*, trvar
predict gainw*, trgain
predict weiw*, trweight

gendir = 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 vv 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)
Example using seqlogit

```
. 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).

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

(Continued on next page)
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)
Example using `seqlogit`

```
. uhdesc, draws(10) at(south 0 country 0 suburb 0 city 0)

<table>
<thead>
<tr>
<th>Transition</th>
<th>p(atrisk)</th>
<th>mean(e)</th>
<th>sd(e)</th>
<th>corr(e,x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transition 1</td>
<td>1.000</td>
<td>-0.000</td>
<td>1.000</td>
<td>-0.000</td>
</tr>
<tr>
<td>Transition 2</td>
<td>0.958</td>
<td>0.040</td>
<td>0.982</td>
<td>-0.026</td>
</tr>
<tr>
<td>Transition 3</td>
<td>0.218</td>
<td>0.607</td>
<td>0.902</td>
<td>-0.135</td>
</tr>
</tbody>
</table>

. uhdesc, overat(coh 1.5, coh 3 , coh 4.5, coh 6, coh 7.5) ///
> draws(10) overlay(1915 1930 1945 1960 1975) ///
> at(south 0 country 0 suburb 0 city 0)

<table>
<thead>
<tr>
<th>Year</th>
<th>Transition 1</th>
<th>p(atrisk)</th>
<th>mean(e)</th>
<th>sd(e)</th>
<th>corr(e,x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1915</td>
<td>Transition1</td>
<td>1.000</td>
<td>-0.000</td>
<td>1.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>Transition2</td>
<td>0.896</td>
<td>0.094</td>
<td>0.963</td>
<td>-0.053</td>
</tr>
<tr>
<td></td>
<td>Transition3</td>
<td>0.196</td>
<td>0.701</td>
<td>0.896</td>
<td>-0.123</td>
</tr>
<tr>
<td>1930</td>
<td>Transition1</td>
<td>1.000</td>
<td>-0.000</td>
<td>1.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>Transition2</td>
<td>0.929</td>
<td>0.065</td>
<td>0.973</td>
<td>-0.039</td>
</tr>
<tr>
<td></td>
<td>Transition3</td>
<td>0.209</td>
<td>0.660</td>
<td>0.899</td>
<td>-0.129</td>
</tr>
<tr>
<td>1945</td>
<td>Transition1</td>
<td>1.000</td>
<td>-0.000</td>
<td>1.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>Transition2</td>
<td>0.953</td>
<td>0.044</td>
<td>0.981</td>
<td>-0.028</td>
</tr>
<tr>
<td></td>
<td>Transition3</td>
<td>0.217</td>
<td>0.618</td>
<td>0.902</td>
<td>-0.134</td>
</tr>
<tr>
<td>1960</td>
<td>Transition1</td>
<td>1.000</td>
<td>-0.000</td>
<td>1.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>Transition2</td>
<td>0.969</td>
<td>0.029</td>
<td>0.987</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td>Transition3</td>
<td>0.220</td>
<td>0.572</td>
<td>0.903</td>
<td>-0.135</td>
</tr>
<tr>
<td>1975</td>
<td>Transition1</td>
<td>1.000</td>
<td>-0.000</td>
<td>1.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>Transition2</td>
<td>0.980</td>
<td>0.019</td>
<td>0.991</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>Transition3</td>
<td>0.215</td>
<td>0.519</td>
<td>0.904</td>
<td>-0.130</td>
</tr>
</tbody>
</table>
```

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

|   | Odds Ratio | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|---|------------|-----------|------|------|----------------------|
| .1_2_3_4v0 |            |           |      |      |                      |
| sibs | .8017449   | .0136828  | -12.95 | .000 | .7753708 .8290162   |
| south | .4585072   | .0439189  | -8.14  | .000 | .3800262 .5531971   |
| country | .8939824   | .0940679  | -1.07  | .287 | .7273822 1.098741   |
| suburb | 1.289316   | .2341737  | 1.40   | .162 | .903148 1.840603    |
| city | .8692876   | .1340587  | 0.91   | .364 | .6425308 1.17607    |
| coh | 1.437848   | .1282836  | 4.25   | .000 | 1.216166 1.699397   |
| paeduc | 1.20536    | .051681  | 4.36   | .000 | 1.108206 1.311031   |
| _paeduc_X_h | .9952772  | .0091567  | -0.51  | .607 | .9774912 1.013387   |

|   | Odds Ratio | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|---|------------|-----------|------|------|----------------------|
| .2v1 |            |           |      |      |                      |
| sibs | .931144    | .0200288  | -3.32 | .001 | .8927042 .971239    |
| south | .8638969   | .0936012  | -1.35 | .177 | .6986122 1.068286   |
| country | 1.051554   | .1233927  | 0.43  | .668 | .8355043 1.32347    |
| suburb | 1.26571    | .1678307  | 1.78  | .076 | .9760373 1.641352   |
| city | .7376786   | .1259632  | -1.78 | .075 | .5278607 1.030897   |
| coh | 1.586091   | .2202713  | 3.32  | .001 | 1.208137 2.052286   |
| paeduc | 1.135127   | .0698685  | 2.06  | .039 | 1.006125 1.280669   |
| _paeduc_X_h | .9916887  | .0064161  | -0.65 | .516 | .975607 1.013387    |

|   | Odds Ratio | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|---|------------|-----------|------|------|----------------------|
| .3_4v1 |            |           |      |      |                      |
| sibs | .8180824   | .0130515  | -12.53 | .000 | .7297595 .844754    |
| south | .9660915   | .0701366  | -0.48 | .638 | .8379583 1.113818   |
| country | 1.117765   | .0918093  | 1.36  | .175 | .9515595 1.313002   |
| suburb | 1.429422   | .1292345  | 3.95  | .000 | 1.1973 1.706645    |
| city | .9884489   | .1030442  | -0.11 | .911 | .8057826 1.215265   |
| coh | .8516002   | .0796486  | -1.72 | .086 | .7089643 1.022933   |
| paeduc | 1.130445   | .0427277  | 3.24  | .001 | 1.049727 1.217369   |
| _paeduc_X_h | 1.002702   | .0076537  | 2.74  | .006 | 1.005838 1.03584    |

|   | Odds Ratio | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|---|------------|-----------|------|------|----------------------|
| .4v3 |            |           |      |      |                      |
| sibs | .9028393   | .0256562  | -3.60 | .000 | .8539288 .954513    |
| south | .8578608   | .060945   | -1.24 | .215 | .6732028 1.09317    |
| country | 1.305278   | .1911589  | 1.82  | .069 | .979887 1.73925    |
| suburb | 1.418216   | .199573   | 2.48  | .013 | 1.076366 1.868634   |
| city | 1.769032   | .293412   | 3.37  | .001 | 1.269704 2.446729   |
| coh | .6962884   | .145539   | -2.20 | .028 | .5043695 .961235   |
| paeduc | .9991627   | .059441   | -0.01 | .989 | .8891961 1.122729   |
| _paeduc_X_h | 1.005108   | .0118746  | 0.43  | .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
Example using seqlogit

```
. uhdesc, draws(10) at(south 0 country 0 suburb 0 city 0)

<table>
<thead>
<tr>
<th></th>
<th>p(atrisk)</th>
<th>mean(e)</th>
<th>sd(e)</th>
<th>corr(e,x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>transition 1</td>
<td>1.000</td>
<td>-0.000</td>
<td>1.000</td>
<td>0.200</td>
</tr>
<tr>
<td>transition 2</td>
<td>0.958</td>
<td>0.040</td>
<td>0.982</td>
<td>0.168</td>
</tr>
<tr>
<td>transition 3</td>
<td>0.218</td>
<td>0.607</td>
<td>0.902</td>
<td>0.037</td>
</tr>
</tbody>
</table>
```

References
