Mixing cross-sectional and longitudinal designs

SEQUENTIAL COHORT DESIGN

Expanding the number of time points

- Repeated measurements are expensive
- Basic simultaneous cross-sectional studies can also provide information on age-related effects
  - Just treat age as time!
  - The key assumption is that there are no cohort effects
- No intra-individual change can be assessed, only group effects
- Useful in educational research

Age group	Sample	Occasion	Variables	Implied occasion
A1	S1	T1	X1, X2,…Xm	T1
A2	S2	T1	X1, X2,…Xm	T2
…	…			
Ag	Sg	T1	X1, X2,…Xm	Tg

Mixing cross-sectional and longitudinal

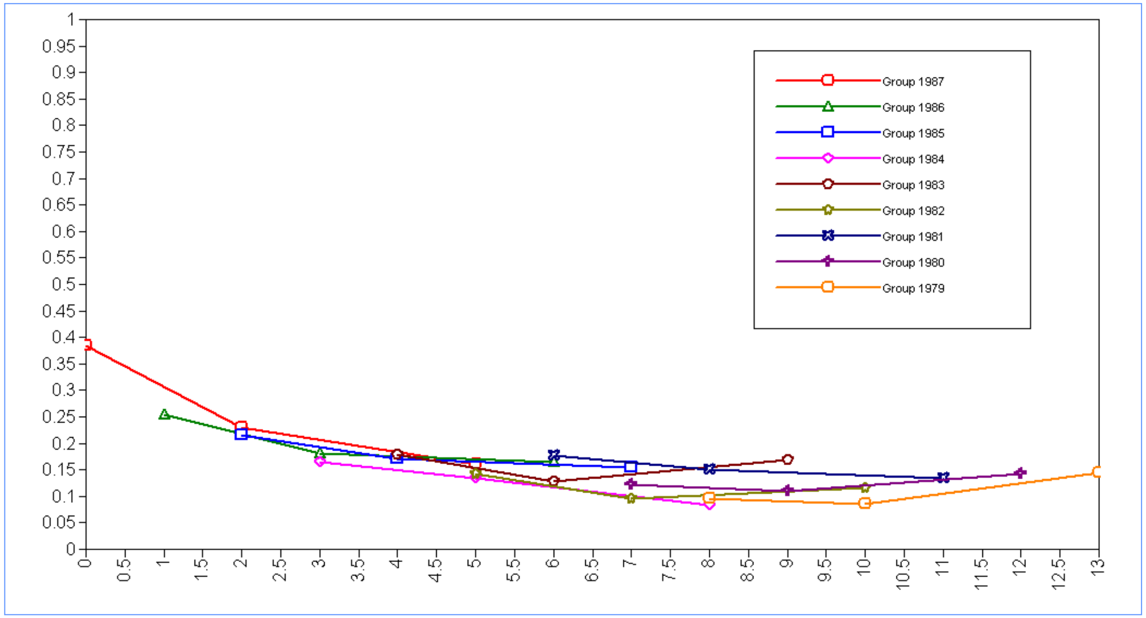
- If there are some repeated measurements, their number can be expanded by treating age as time
- For example, if age groups are one year apart, and the measurement occasions are one year apart, the following treatment of data is possible:

Age group	Sample	Occasion	Variables	Implied occasions
A1	S1	T1, T2, T3	X1, X2,…Xm	T1,T2,T3
A2	S2	T1, T2, T3	X1, X2,…Xm	T2,T3,T4
…	…			
Ag	Sg	T1, T2, T3	X1, X2,…Xm	T_{g-2}, T_{g-1}, Tg

- Improvement on the cross-sectional design, as the assumption of equivalence of cohorts can be tested

Sequential cohort design

- Latent Growth Cohort-Sequential (or accelerated) design links adjacent segments of repeated data from different age cohorts to estimate a common developmental trend or growth curve
  - Each cohort has a different pattern of "missingness"
  - It is possible to build the complete curve using information from all cohorts simultaneously



Study of drinking habits in young people

- Research question: Development of alcohol use from age 16 to 29
- Sample: community sample of Swiss urban adolescents and young adults aged 16 to 24 (N=2840)
- Occasions: baseline 2003; 2-year follow up, 5-year follow up
- Measure: Frequency of alcohol use during the month prior to the interviews using 5 response categories: 0=never, 1=1–3 times a month, 2=1–2 times a week, 3=3–6 times a week, 4=daily.

Age at measurement occasions

9 cohorts
3 repeated measurements

Cohort	2003	2005	2008
1987	16	18	21
1986	17	19	22
1985	18	20	23
1984	19	21	24
1983	20	22	25
1982	21	23	26
1981	22	24	27
1980	23	25	28
1979	24	26	29

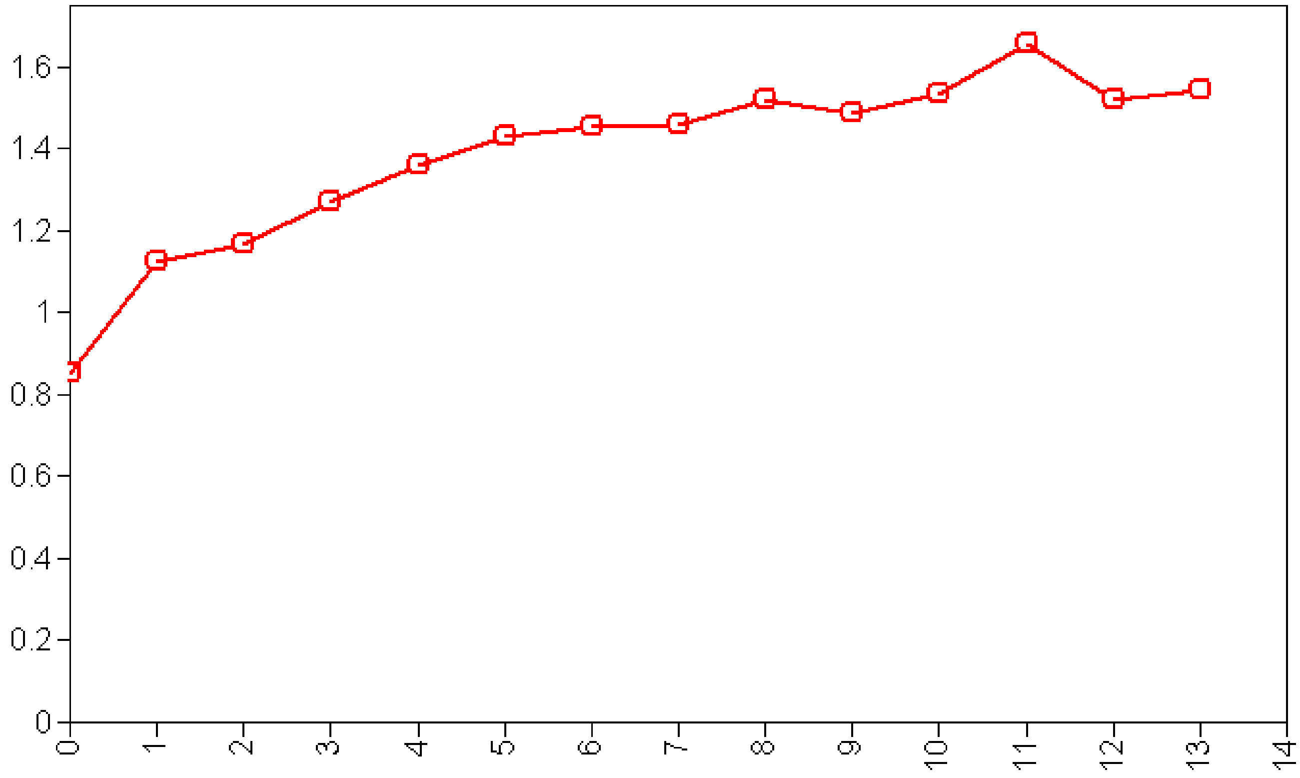
Age as time

Age >> Cohort	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1987	t1		t2			t3								
1986		t1		t2			t3							
1985			t1		t2			t3						
1984				t1		t2			t3					
1983					t1		t2			t3				
1982						t1		t2			t3			
1981							t1		t2			t3		
1980								t1		t2			t3	
1979									t1		t2			t3
Time score	0	1	2	3	4	5	6	7	8	9	10	11	12	13

Data mapping approach

- DATA COHORT syntax option in Mplus – works out the time score based on birth year and measurement year
- Only works with continuous variables!
- Let's pretend that our "alcohol use" variables are continuous and check out this option
- The idea is to re-map our cohort and occasion variables as new time score
- Then specify a growth model for the whole time span (14 years)
  - Let's hypothesise a quadratic growth curve
  - Drinking will steadily increase, reach a pick in mid 20th, and then decrease

Observed means



PLOT: TYPE IS PLOT3;
SERIES = t1alk t2alk t3alk (slp);

Accelerated cohort syntax

VARIABLE: !some other commands here

DATA COHORT:

COHORT IS BirthY (1987 1986 1985 1984 1983 1982 1981 1980 1979);

TIMEMEASURES= t1alk (2003) t2alk (2005) t3alk (2008);

TNAMES = alk;

Centring on the middle time point is often better for quadratic curves

MODEL:

int slope qu | alk16@-.7 alk17@-.6 alk18@-.5 alk19@-.4 alk20@-.3 alk21@-.2 alk22@-.1 alk23@0 alk24@.1

alk25@.2 alk26@.3 alk27@.4 alk28@.5 alk29@.6;

alk16-alk29* (1); !assume residual variances the same across time

Results with continuous data: fit

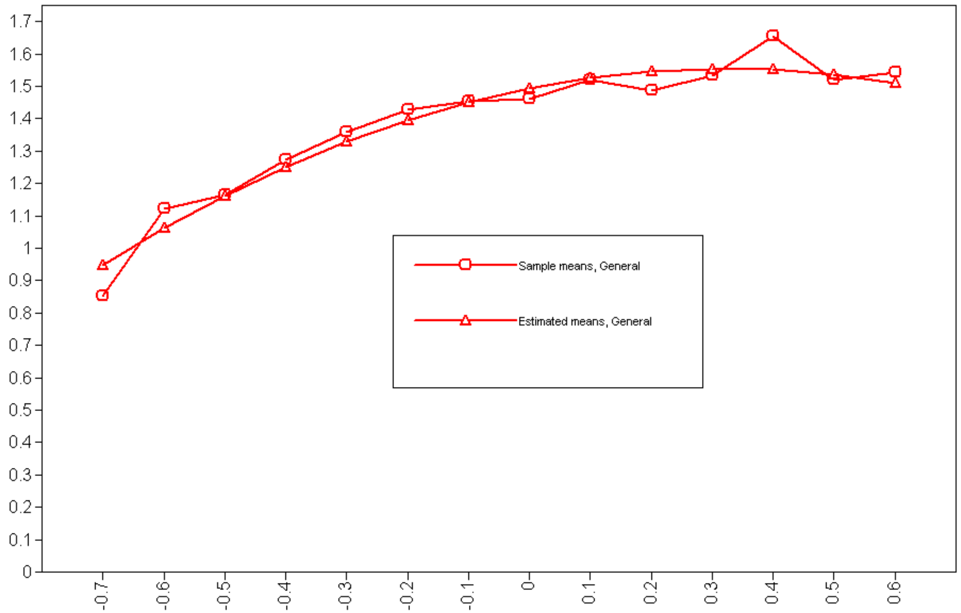
- Model fit is not great but not too bad either

Chi-Square Test of Model Fit		
	Value	100.968
	Degrees of Freedom	45
	P-Value	0.0000
CFI		0.968
TLI		0.981
RMSEA (Root Mean Square Error Of Approximation)		
	Estimate	0.021
	90 Percent C.I.	0.015 0.026

Model results

	Estimate	S.E.	Est./S.E.	P-Value
Means				
INT	1.493	0.015	97.158	0.000
SLOPE	0.374	0.031	12.159	0.000
QU	-0.575	0.065	-8.868	0.000
Variances				
INT	0.433	0.019	22.958	0.000
SLOPE	0.473	0.131	3.606	0.000
QU	1.193	0.367	3.252	0.001
SLOPE WITH				
INT	0.058	0.026	2.255	0.024
QU WITH				
INT	-0.426	0.070	-6.084	0.000
SLOPE	0.197	0.095	2.069	0.039
Residual Variances				
ALK16	0.358	0.009	38.193	0.000
ALK17	0.358	0.009	38.193	0.000

etc……



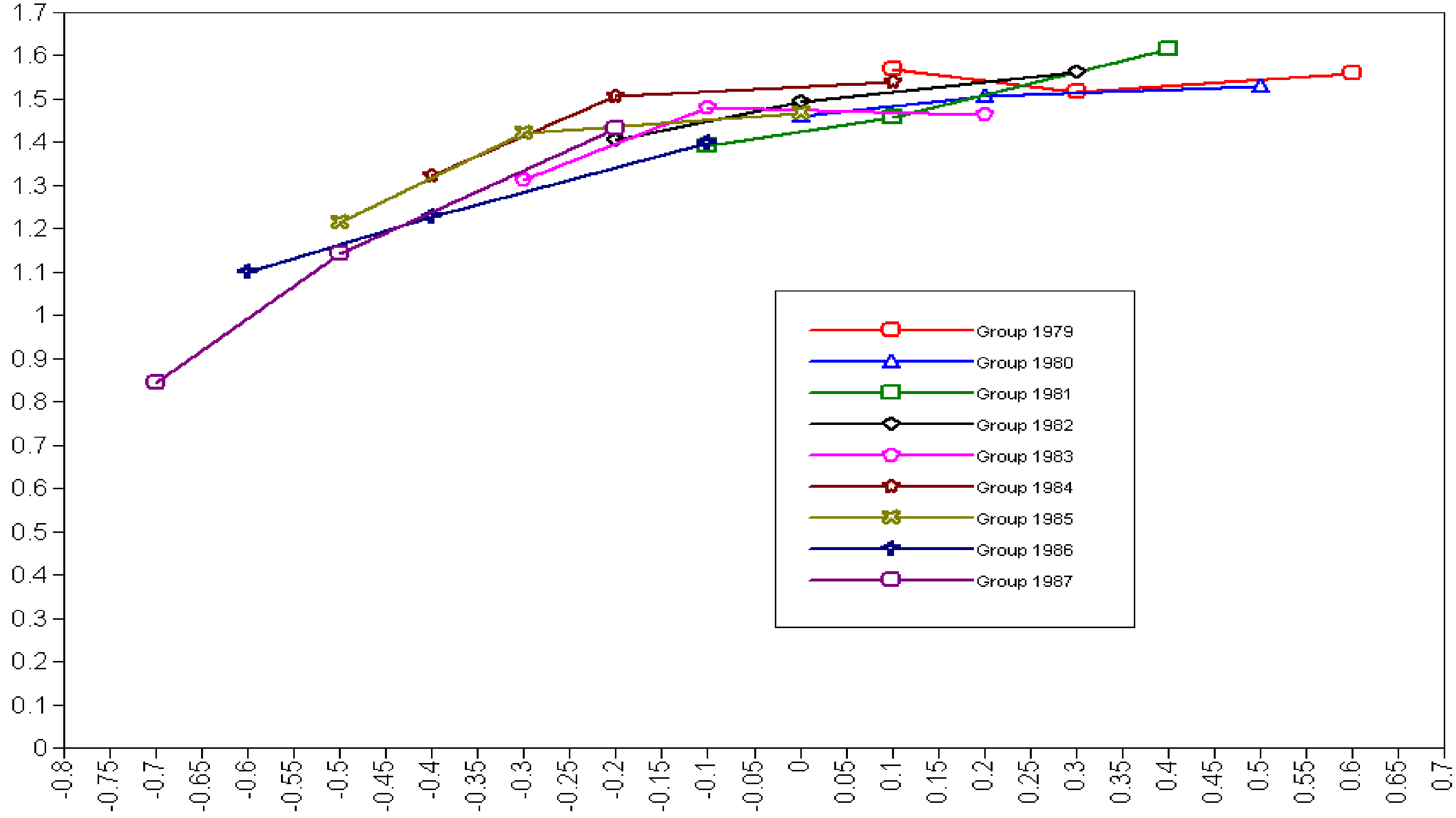
Discussion of the DATA COHORT approach

- Even if no data is missing due to nonresponse, there is plenty of missing data by design
  - Each individual only has 3 non-missing responses, and 11 missing responses
  - can be considered MCAR because these responses were never collected
- However, this approach assumes that we actually had 14 data collection occasions
  - Which we did not
  - Are the degrees of freedom correct?

Multi-group approach

- The idea is to specify a growth model for each of the cohorts (using the new time score)
- And then test if the same model holds for all cohorts
- Different cohorts will have different occasions present
  - Missing by design (MCAR)
- Treat cohorts as multiple groups with their own measurement occasions
- Importantly, to maintain common growth model, its parameters have to be constrained equal across cohorts

Observed means



** for comparability of results, we pretend that data is continuous

Sequential cohort multi-group syntax: common model

```
MODEL:
 ! This is the common model, and also model for the 1987 cohort
 INT SLP QU | t1alk@-.7 t2alk@-.5 t3alk@-.2;

 !These constraints mean that the samples are drawn from the same population
 INT (1);      !variance of the intercept is the same across samples
 SLP (2);      !variance of the slope is the same
 QU (3);       !variance of the quadratic term is the same
 [INT] (4);    !mean of the intercept is the same
 [SLP] (5);    !mean of the slope is the same
 [QU] (6);     !mean of the quadratic term is the same
 INT WITH SLP*0 (7);   !and all covariances are the same
 INT WITH QU*0 (8);
 SLP WITH QU*0 (9);

 t1alk-t3alk* (10);    !residuals are assumed equal across time
```

Same middle-point centring as before

Sequential cohort multi-group syntax: cohort-specific models

```
MODEL 1986:  INT SLP QU| t1alk@-.6  t2alk@-.4   t3alk@-.1;
MODEL 1985:  INT SLP QU| t1alk@-.5  t2alk@-.3   t3alk@0;
MODEL 1984:  INT SLP QU| t1alk@-.4  t2alk@-.2   t3alk@.1;
MODEL 1983:  INT SLP QU| t1alk@-.3  t2alk@-.1   t3alk@.2;
MODEL 1982:  INT SLP QU| t1alk@-.2  t2alk@0     t3alk@.3;
MODEL 1981:  INT SLP QU| t1alk@-.1  t2alk@.1    t3alk@.4;
MODEL 1980:  INT SLP QU| t1alk@0    t2alk@.2    t3alk@.5;
MODEL 1979:  INT SLP QU| t1alk@.1   t2alk@.3    t3alk@.6;
```

Model results: exact fit

- Degrees of freedom differ from the DATA COHORT approach

Chi-Square Test of Model Fit

Value	142.521
Degrees of Freedom	71
P-Value	0.0000

Notice difference in DF from slide 11 (was 45)

- Now we can see chi-square contributions from each group

1979	19.793	
1980	13.282	
1981	10.609	smallest
1982	26.590	largest
1983	11.726	
1984	14.958	
1985	13.289	
1986	15.708	
1987	16.566	

Model results: approximate fit

- Fit indices are a little worse than in the DATA COHORT approach

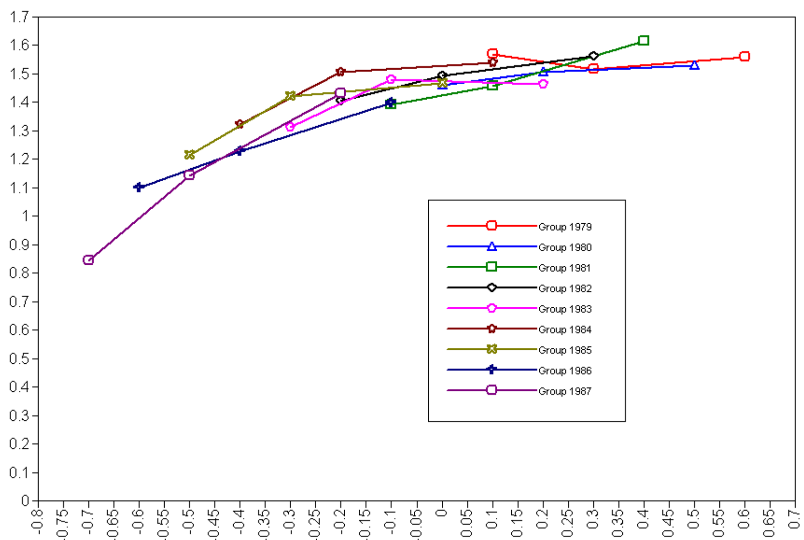
RMSEA (Root Mean Square Error Of Approximation)		
Estimate	0.057	
90 Percent C.I.	0.043	0.070
Probability RMSEA <= .05		0.204
CFI	0.959	
TLI	0.984	

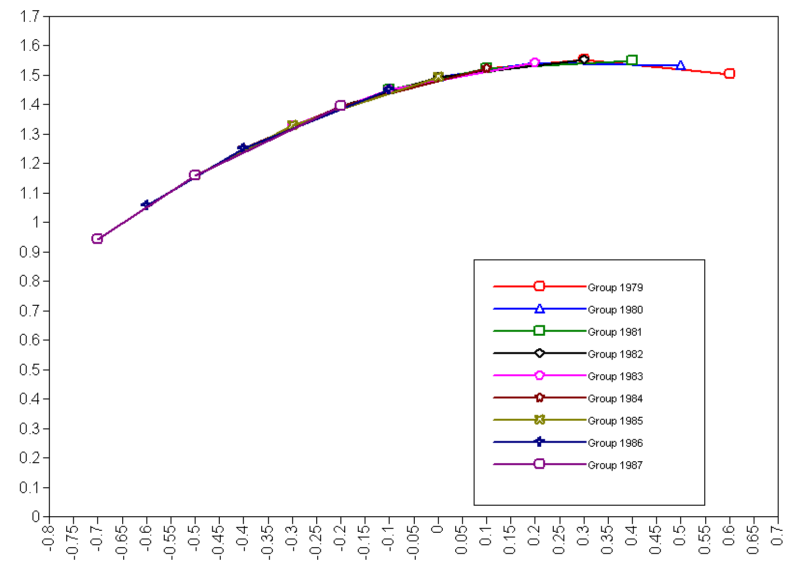
Model results: means

–	Means	Estimate	S.E.	Est./S.E.	P-Value
–	INT	1.493	0.015	97.160	0.000
–	SLP	0.374	0.031	12.161	0.000
–	QU	-0.575	0.065	-8.866	0.000

Means are exactly the same as in the DATA COHORT model (slide 12)

- Observed and estimated means plotted





Model results: variance

- Variances
- INT 0.433 0.019 22.958 0.000
- SLP 0.473 0.131 3.605 0.000
- QU 1.192 0.367 3.252 0.001

- INT WITH
- SLP 0.058 0.026 2.254 0.024
- QU -0.426 0.070 -6.084 0.000
- SLP WITH
- QU 0.197 0.095 2.070 0.038
-
- Residual Variances
- T1ALK 0.358 0.009 38.193 0.000
- T2ALK 0.358 0.009 38.193 0.000
- T3ALK 0.358 0.009 38.193 0.000

Variances and covariances are also exactly the same (slide 12)

Let's stop pretending

- Having established that the multi-group design works well, we can now consider the ordinal nature of our data
- How often do you drink alcohol?

  0 = never

  1 = 1-3 times a month (party?)

  2 = 1-2 times a week (weekend?)

  3 = 3-6 times a week

  4 = daily

We will collapse the last 2 categories because "daily" is not used in one cohort

- Clearly, increase between these categories is not at the interval level

Changes to accommodate categorical data

- First, declare variables as ordinal

  CATEGORICAL = t1alk t2alk t3alk;
- Next, change estimator

  ESTIMATOR=WLSMV;

  PARAMETERIZATION=THETA; !to constrain residuals
- Categorical variables have no scale.
  - To set the scale, Mplus will automatically fix the mean of our growth intercept to 0, and the residual variance of t1alk to 1. It will do so in the first group only (cohort 1979).
  - We will override these defaults. Since we assume parameters equal across groups, we set the intercept mean to 1.493, and its variance to 0.433 for all groups.
    - We pick the values established in the continuous model just for the fun of it, we could pick any other values.

Syntax for categorical variables: common model

```
MODEL:   ! This is the common model, and also model for the 1987 cohort
 INT SLP QU | t1alk@-.7 t2alk@-.5 t3alk@-.2;

 !The samples are drawn from the same population
 INT@.433;      !variance of the intercept is fixed to set the scale
 SLP  (2);      !variance of the slope is the same across samples
 QU  (3);       !variance of the quadratic term is the same
 [INT@1.493];   !mean of intercept is fixed to set the scale
 [SLP]  (5);    !mean of slope is the same
 [QU]  (6);     !mean of quadratic term is the same
 INT WITH SLP*0  (7);
 INT WITH QU*0  (8);
 SLP WITH QU*0  (9);
 t1alk-t3alk*  (10);    !residual variances are the same across time
```

Syntax for categorical variables: individual cohorts

```
MODEL 1986:  INT SLP QU| t1alk@-.6  t2alk@-.4   t3alk@-.1;
MODEL 1985:  INT SLP QU| t1alk@-.5  t2alk@-.3   t3alk@0;
MODEL 1984:  INT SLP QU| t1alk@-.4  t2alk@-.2   t3alk@.1;
MODEL 1983:  INT SLP QU| t1alk@-.3  t2alk@-.1   t3alk@.2;
MODEL 1982:  INT SLP QU| t1alk@-.2  t2alk@0     t3alk@.3;
MODEL 1981:  INT SLP QU| t1alk@-.1  t2alk@.1    t3alk@.4;
MODEL 1980:  INT SLP QU| t1alk@0    t2alk@.2    t3alk@.5;
MODEL 1979:  INT SLP QU| t1alk@.1   t2alk@.3    t3alk@.6;
!this model will be the first cohort according to Mplus, we need to override defaults
[INT@1.493];
t1alk-t3alk* (10);
```

Sequential cohorts with categorical variables: exact fit

Chi-Square Test of Model Fit

Value	175.615*
Degrees of Freedom	97
P-Value	0.0000

Chi-Square Contributions From Each Group

1979	18.100
1980	11.635
1981	12.920
1982	35.692
1983	14.021
1984	30.366
1985	13.305
1986	17.748
1987	21.828

Sequential cohorts with categorical variables: approximate fit

- Fit indices indicate that fit is better than when using the continuous model

```
RMSEA (Root Mean Square Error Of Approximation)
          Estimate                          0.051
          90 Percent C.I.                      0.039  0.063
          Probability RMSEA <= .05             0.447

CFI/TLI
          CFI                          0.979
          TLI                          0.994
```

Model with categorical variables: results

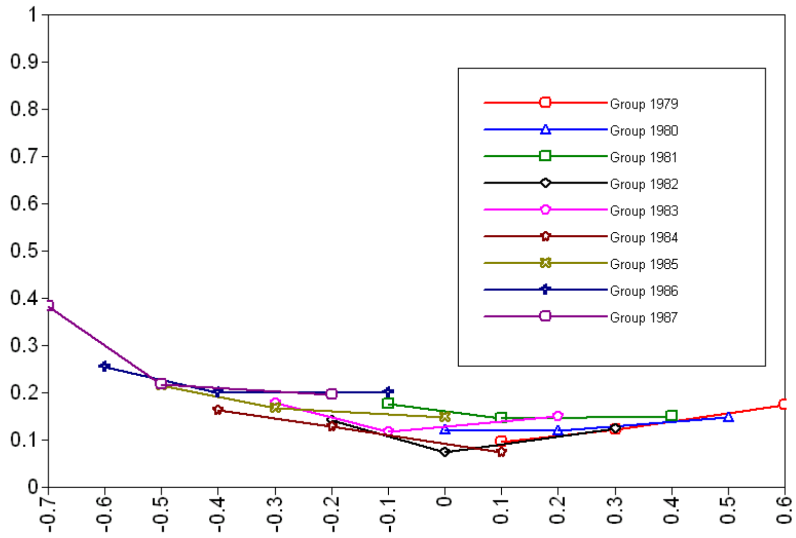
	Estimate	S.E.	Est./S.E.	P-Value
Means				
INT	1.493	0.000	999.000	999.000
SLP	0.363	0.038	9.618	0.000
QU	-0.611	0.075	-8.189	0.000
Variances				
INT	0.433	0.000	999.000	999.000
SLP	0.414	0.130	3.190	0.001
QU	1.391	0.398	3.496	0.000
INT WITH				
SLP	0.063	0.029	2.194	0.028
QU	-0.378	0.064	-5.874	0.000
SLP WITH				
QU	0.225	0.109	2.062	0.039
Residual Variances				
T1ALK	0.261	0.016	16.191	0.000
T2ALK	0.261	0.016	16.191	0.000
T3ALK	0.261	0.016	16.191	0.000

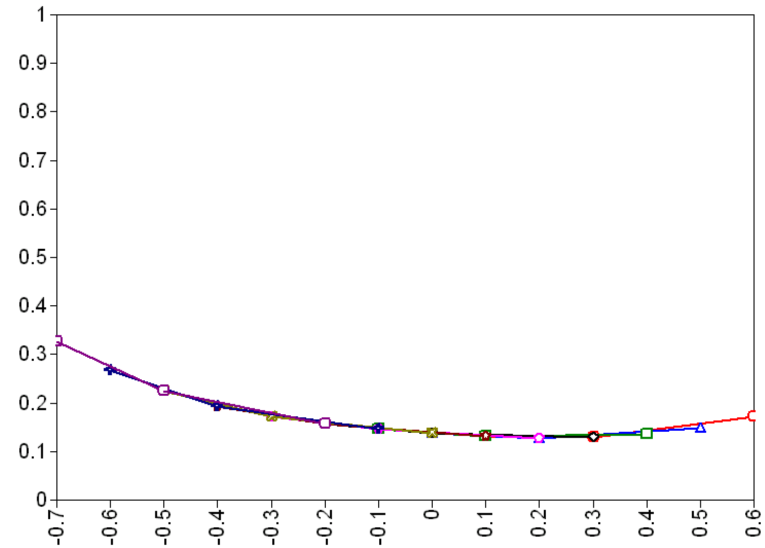
Means are similar to the model with continuous variables (slide 20)

Variances and covariances are also similar (slide 21)

Plots and interpretation

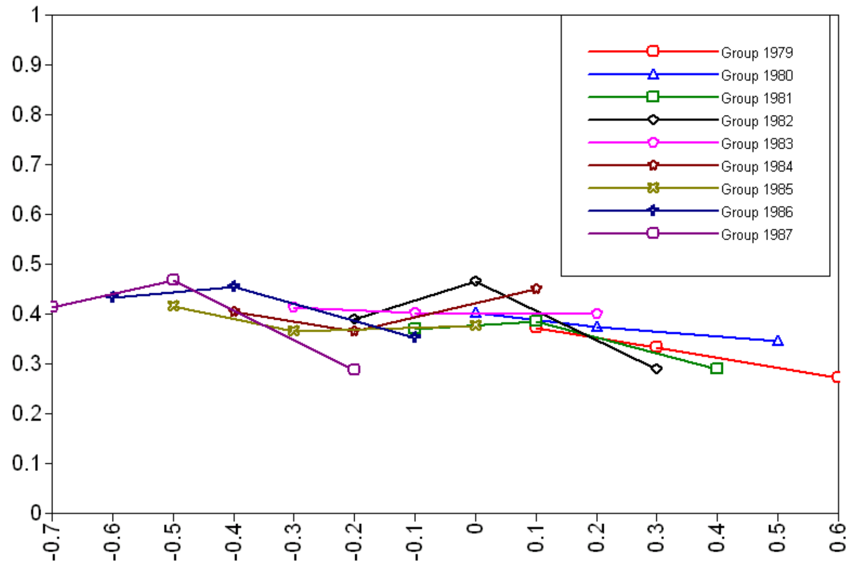
- Plots with categorical data are harder to interpret
- No plots of means, but plots of proportions for a response category
- Here are observed and estimated proportions for the 1st category ("never")
- About 30% of 16 year-olds never drink alcohol, and at the age of 25 this percentage is at its lowest, about 15%

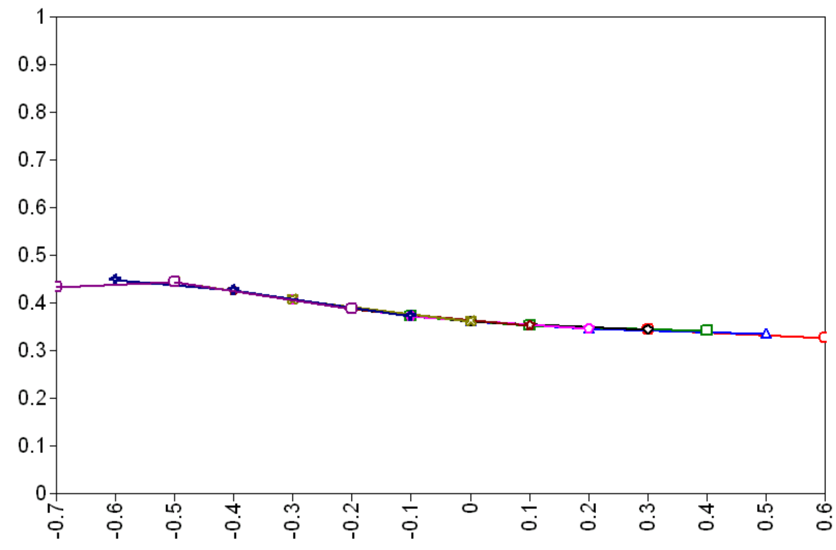




Plots and interpretation – cont.

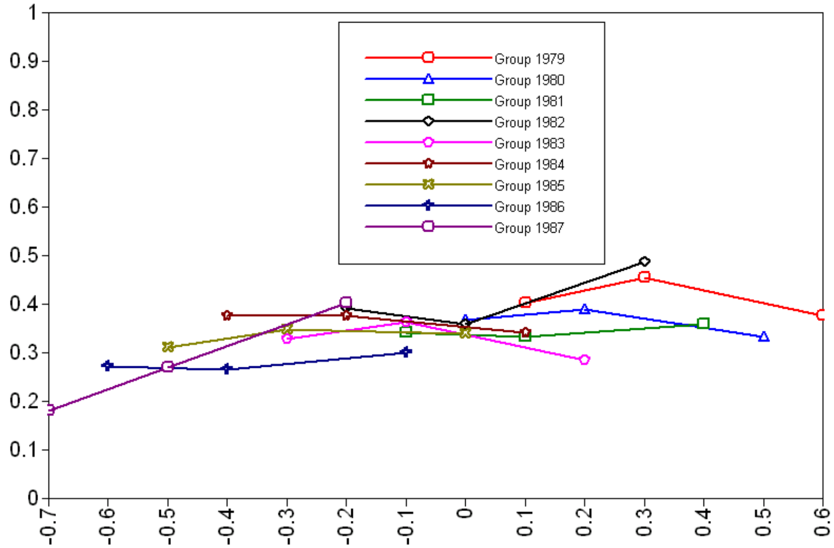
- Here are proportions for the second category ("once a month")
- Between 35% and 45% of young adults drink alcohol once a month

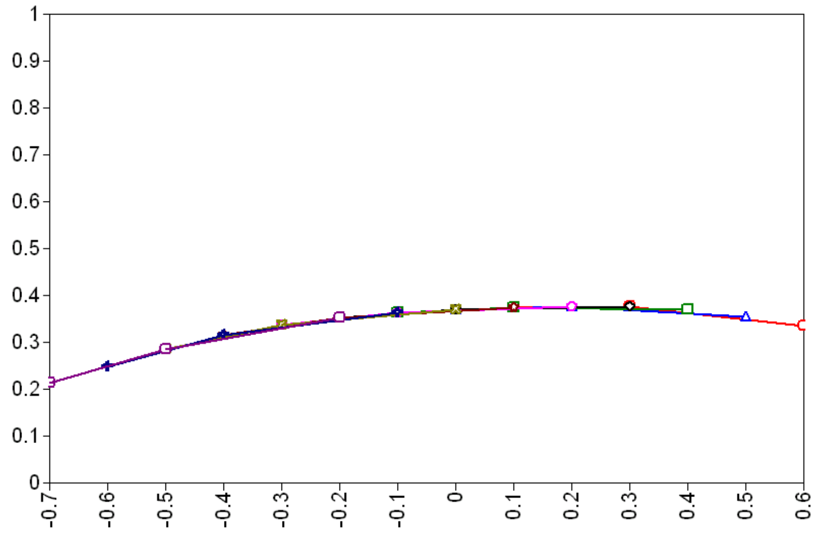




Plots and interpretation – cont.

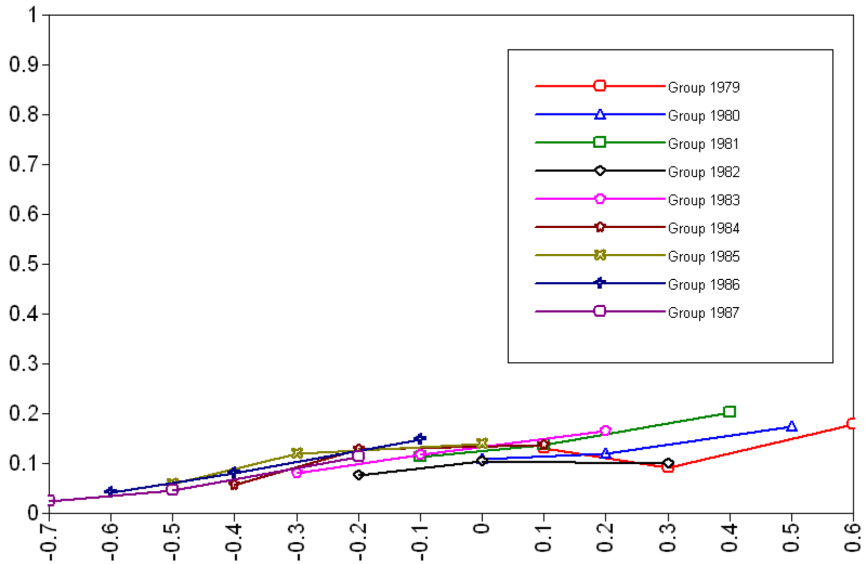
- Here are proportions for the third category ("once a week")
- Between 20% and 45% of young adults drink alcohol once a week

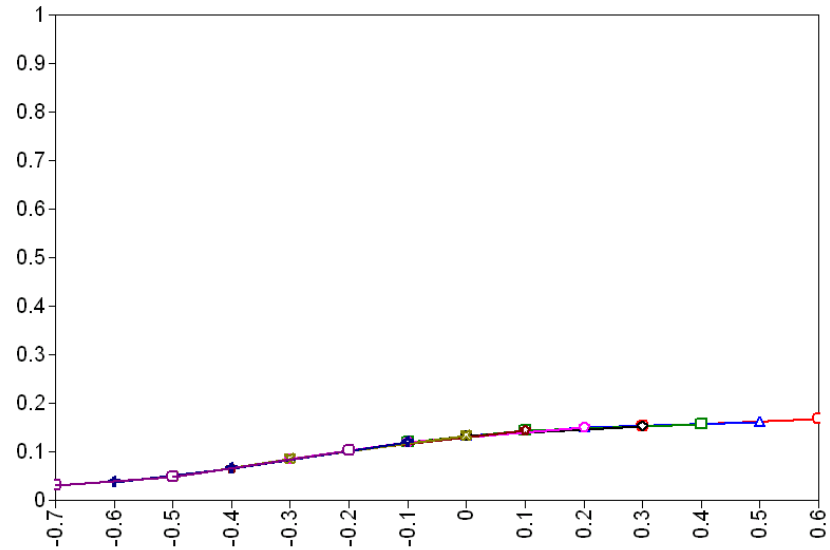




Plots and interpretation – cont.

- Here are proportions for the last category ("3-7 times a week")
- Only about 3% of 16 year-olds drink alcohol as often as this, and by the age of 29 the proportion goes up to about 15%





Testing assumptions

- Our model assumed that all cohorts are from the same population, i.e. there are no cohort effects
  - Means of growth factors are the same
  - Variances and covariances of growth factors are the same
- Mplus “helps” by imposing additional assumptions
  - Measurement invariance (notice that the item thresholds are exactly the same across cohorts)
- We can test whether these assumptions hold
  - Looking at MI, it seems that the youngest cohort has different thresholds at T1, different means of linear and quadratic terms