Annex: problems with APC-IE and HAPC

## This annex shows how both the APC-IE ([Yang et al. 2008](#_ENREF_2)) and the HAPC ([Reither et al. 2009](#_ENREF_1)) document age effects that cannot be true. What do we infer this from?

## It is impossible to lose one’s educational credentials over time, e.g. one cannot lose years of schooling or a Bachelor’s degree as an effect of age. Yet, this is what both the APC-IE and HAPC indicate.

To see this result, use the database of the Luxembourg Income Study ([http://www.lisdatacenter.org](http://www.lisdatacenter.org/)). Note: You need a “Lissy” Account to use the remote access system of the Luxembourg Income Study. Such an account can be created easily. Once you have a login and password, enter Lissy, specify that you want to work in the “LIS” database and enter a title for your project. For the first example on APC-IE, specify that you want to use Stata code; for the second example, where we use the HAPC, specify SAS.

Using the Stata code below with LISSY, we use the US datasets 1986-2010, age groups 25-74, and we look at level of education, which is coded in categories 1, 2 and 3 (see <http://www.lisdatacenter.org/wp-content/uploads/standardisation-of-education-levels.pdf>). We consider a logit-type APC specification, similar to papers that use the APC-IE and HAPC models. To replicate our results, enter the following code into the LISSY remote access system (you can copy and paste this code, but pay attention to page breaks, which you can avoid when viewing this file using Word’s “web layout”):

\*SYNTAX

clear all

foreach gogo in us {

qui {

if "`gogo'"== "us" local fifi "us86 us91 us94 us00 us04 us10"

foreach toto in `fifi' {

local perso "$`toto'p"

local house "$`toto'h"

qui use hid ppopwgt age relation educ using `perso' , clear

qui joinby hid using `house' //Now we merge the variables that we retained from the personal dataset to the entire household dataset.

keep hid ppopwgt age relation educ year iso2 hpopwgt

local save "t`toto'" //Puts into a variable, of which the name is "save" the string t`toto', e.g. the dataset us10.dta with the variables that we want.

qui save `save' , replace //Before we just defined the local, now we actually safe.

} //Now we close the loop, for each country we have five different years with all the info that we need, but we still need to merge each country and year into one file.

clear

foreach toto in `fifi' { //We have five separate files per country and they must be aggragated into a single one

local save "t`toto'" //now it safes e.g. "us00" "us10" etc. up to the point where every files are for one country, we are still and will still be till the end looking at one country

qui append using `save' //"save" is simply the file name to which everything is safed

}

keep if relat <3000 // we retain individuals when independant from parents

qui recode year (1968/1971=1970) (1972/1976=1975) (1977/1982=1980) (1983/1987=1985) (1988/1992=1990) (1993/1997=1995) (1998/2002=2000) (2003/2006=2005) (2007/2012=2010) //We recode the years into five-year intervals

\* Generates the weights that we need

qui gen pweight = int(ppop)

\*Here we assign everyone to a five-year interval, always to the lowest one 26=25 and 29=25

gen page=floor(age/5)\*5

\*Here we assign everything to a five-year interval, so year is in five-year intervals and year5 is the original variable

gen year5=year

replace year =int((year-1980)/5)

gen coco = year5 - page // cohort

}

di "\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"

di "COUNTRY NAME: `gogo'" //this simply tells us the name of the country in the listing

keep if year5<=2010 & year5>=1985 & page>=25 & page<75

su \_all

gen seduc = educ==3

xi: apc\_ie seduc [pw= pweight] if year5<=2010 & year5>=1985 & page>=25 & page<70, f(bin) l(logit) age(page) period(year5)

}

This code results in the following listing from LISSY. These results indicate that people lose their educational credentials with age starting at age 30 (the indicators for “age\_xx” decline after the indicator “age\_30”).

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

COUNTRY NAME: us

(63216 observations deleted)

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

hid | 454991 32104.02 20238.92 1 76447

ppopwgt | 454991 1862.422 1657.179 .01 44118.2

relation | 454991 1413.282 527.8846 1000 2200

age | 454991 46.23227 13.00036 25 74

educ | 454991 2.218163 .674143 1 3

-------------+--------------------------------------------------------

iso2 | 0

year | 454991 4.062467 1.538635 1 6

hpopwgt | 454991 1861.441 1656.349 .01 44118.2

pweight | 454991 1862.012 1657.179 0 44118

page | 454991 44.24945 12.98409 25 70

-------------+--------------------------------------------------------

year5 | 454991 2000.312 7.693176 1985 2010

coco | 454991 1956.063 14.90308 1915 1985

Iteration 0: log pseudolikelihood = -5.060e+08

Iteration 1: log pseudolikelihood = -5.053e+08

Iteration 2: log pseudolikelihood = -5.053e+08

Iteration 3: log pseudolikelihood = -5.053e+08

Intrinsic estimator of APC effects No. of obs = 431198

Optimization : ML Residual df = 431172

Scale parameter = 1

Deviance = 1010668886 (1/df) Deviance = 2344.004

Pearson = 800317728.6 (1/df) Pearson = 1856.145

Variance function: V(u) = u\*(1-u) [Bernoulli]

Link function : g(u) = ln(u/(1-u)) [Logit]

AIC = 2343.863

Log pseudolikelihood = -505334443.1 BIC = 1.01e+09

------------------------------------------------------------------------------

| Robust

seduc | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

age\_25 | .0672781 .0180665 3.72 0.000 .0318684 .1026878

age\_30 | .1946057 .0138178 14.08 0.000 .1675233 .2216881

age\_35 | .1805522 .0122205 14.77 0.000 .1566004 .2045039

age\_40 | .1661336 .0114745 14.48 0.000 .1436439 .1886233

age\_45 | .1136776 .0115217 9.87 0.000 .0910954 .1362597

age\_50 | .0025228 .012796 0.20 0.844 -.022557 .0276026

age\_55 | -.1221157 .0146136 -8.36 0.000 -.1507578 -.0934736

age\_60 | -.2314502 .0164139 -14.10 0.000 -.2636207 -.1992796

age\_65 | -.3712041 .0174817 -21.23 0.000 -.4054675 -.3369406

period\_1985 | -.5887601 .0183963 -32.00 0.000 -.6248162 -.5527039

period\_1990 | -.175567 .0102174 -17.18 0.000 -.1955927 -.1555412

period\_1995 | -.0321985 .0094119 -3.42 0.001 -.0506455 -.0137516

period\_2000 | .1370267 .0091124 15.04 0.000 .1191668 .1548865

period\_2005 | .2418947 .0084655 28.57 0.000 .2253026 .2584867

period\_2010 | .4176043 .0101209 41.26 0.000 .3977677 .4374408

cohort\_1920 | -.1710096 .0779741 -2.19 0.028 -.3238361 -.0181831

cohort\_1925 | -.2159025 .044123 -4.89 0.000 -.302382 -.1294229

cohort\_1930 | -.2039455 .0342913 -5.95 0.000 -.2711551 -.1367359

cohort\_1935 | -.1395654 .0290644 -4.80 0.000 -.1965307 -.0826002

cohort\_1940 | -.0580345 .023978 -2.42 0.016 -.1050304 -.0110385

cohort\_1945 | .1224186 .0192164 6.37 0.000 .0847552 .160082

cohort\_1950 | .2338563 .0169116 13.83 0.000 .2007102 .2670023

cohort\_1955 | .1427318 .0150015 9.51 0.000 .1133294 .1721343

cohort\_1960 | .0014827 .0131751 0.11 0.910 -.0243401 .0273055

cohort\_1965 | .0196264 .0107431 1.83 0.068 -.0014298 .0406826

cohort\_1970 | .0638769 .0117046 5.46 0.000 .0409363 .0868175

cohort\_1975 | .0809791 .0136955 5.91 0.000 .0541363 .1078218

cohort\_1980 | .0433656 .0169914 2.55 0.011 .0100631 .0766681

cohort\_1985 | .0801201 .0295379 2.71 0.007 .0222268 .1380133

\_cons | -.7518052 .0092539 -81.24 0.000 -.7699424 -.7336679

------------------------------------------------------------------------------

.

.

.

end of do-file

As one can see, the coefficients for age, starting with the parameter “age\_30” decrease to lower and lower numbers. This means that, according to this model, people lose their educational credentials with age. This is impossible, since one cannot lose a given level of education with age.

The HAPCD produces similarly faulty estimations. To understand these, again use LISSY, but set it to code for the software package SAS. Then enter the following code as a job text (again, make sure that there are no page breaks):

OPTIONS NONOTES NOSOURCE NOFMTERR NODATE NONUMBER NOCENTER LABEL LS=MAX PS=MAX ;

TITLE "";

%MACRO sample ;

%LET i = 1 ;

%DO %UNTIL (&i > 6) ;

%LET ccyy = %SCAN(&all,&i) ;

TITLE "Country : &ccyy" ;

DATA &ccyy.h ;

SET &&&ccyy.h (KEEP=hid own) ;

RUN ;

PROC SORT DATA=&ccyy.h ;

BY hid ;

RUN ;

DATA &ccyy.p ;

SET &&&ccyy.p (KEEP= hid year dname pwgt ppopwgt relation partner children age sex immigr

educ

educ\_c emp status1 ptime gross1);

RUN ;

PROC SORT DATA=&ccyy.p ;

BY hid ;

RUN ;

DATA &ccyy ;

MERGE &ccyy.h &ccyy.p ;

BY hid ;

IF ((25 <= age <= 74) AND (relation LE 2200)) ;

IF (100 <= own <= 199) THEN homeowner = 1;

IF (200 <= own <= 299) THEN homeowner = 0;

RUN ;

%LET i = %EVAL(&i+1) ;

%END ;

%MEND sample ;

%LET all = us86 us91 us97 us00 us04 us10 ;

%sample

;

DATA ttt;

SET us86 us91 us97 us00 us04 us10 ;

IF (year=1986) then year=1985 ;

IF (year=1991) then year=1990 ;

IF (year=1997) then year=1995 ;

IF (year=2004) then year=2005 ;

ag5 = 5\*INT(age/5) ;

co5 = year- ag5 ;

run;

proc freq data=ttt ;

tables ag5 ;

run;

DATA ttt;

SET ttt ;

IF (ag5=25) then ag5=-4.5;

IF (ag5=30) then ag5=-3.5;

IF (ag5=35) then ag5=-2.5;

IF (ag5=40) then ag5=-1.5;

IF (ag5=45) then ag5=-.5;

IF (ag5=50) then ag5=.5;

IF (ag5=55) then ag5=1.5;

IF (ag5=60) then ag5=2.5;

IF (ag5=65) then ag5=3.5;

IF (ag5=70) then ag5=4.5;

ag52=ag5\*ag5;

RUN;

proc freq data=ttt ;

tables ag5 ;

run;

proc glimmix data=ttt maxopt=25000;

class year co5;

model educ(event='3') = ag5 ag52 /solution

dist=gamma link=log;

random co5 year /solution;

NLOPTIONS TECHNIQUE=NRRIDG;

output out=out1 pred=pred1;

title "Educ: HAPC-CCREM of university educ, CPS 1985-2010";

run;

With these specifications, LISSY gives the following results:

~------------------------------ NOTICE TO USERS ------------------------------~

~ ~

~ Use of the data in the LUXEMBOURG INCOME STUDY DATABASE is governed by ~

~ regulations which do not allow copying or further distribution of the ~

~ survey microdata. ~

~ ~

~ Anyone violating these regulations will lose all privileges to the ~

~ databases and may be subject to prosecution under the law. In addition, ~

~ any attempt to circumvent the LIS processing system or unauthorized entry ~

~ into the LIS computing system will result in prosecution. ~

~ ~

~ All papers written using the LUXEMBOURG INCOME STUDY DATABASE must be ~

~ submitted for entry into the Working Papers Series. ~

~ Please consult our web site for more information at WWW.LISDATACENTER.ORG ~

~ ~

~------------------------------ NOTICE TO USERS ------------------------------~

Country : us10

The FREQ Procedure

Cumulative Cumulative

ag5 Frequency Percent Frequency Percent

--------------------------------------------------------

25 42932 9.60 42932 9.60

30 55176 12.34 98108 21.95

35 59643 13.34 157751 35.29

40 61468 13.75 219219 49.04

45 56210 12.57 275429 61.61

50 48541 10.86 323970 72.47

55 39377 8.81 363347 81.28

60 33125 7.41 396472 88.69

65 27597 6.17 424069 94.87

70 22952 5.13 447021 100.00

Country : us10

The FREQ Procedure

Cumulative Cumulative

ag5 Frequency Percent Frequency Percent

---------------------------------------------------------

-4.5 42932 9.60 42932 9.60

-3.5 55176 12.34 98108 21.95

-2.5 59643 13.34 157751 35.29

-1.5 61468 13.75 219219 49.04

-0.5 56210 12.57 275429 61.61

0.5 48541 10.86 323970 72.47

1.5 39377 8.81 363347 81.28

2.5 33125 7.41 396472 88.69

3.5 27597 6.17 424069 94.87

4.5 22952 5.13 447021 100.00

Educ: HAPC-CCREM of university educ, CPS 1985-2010

The GLIMMIX Procedure

Model Information

Data Set WORK.TTT

Response Variable educ

Response Distribution Gamma

Link Function Log

Variance Function Default

Variance Matrix Not blocked

Estimation Technique Residual PL

Degrees of Freedom Method Containment

Class Level Information

Class Levels Values

year 6 1985 1990 1995 2000 2005 2010

co5 15 1915 1920 1925 1930 1935 1940 1945 1950 1955

1960 1965 1970 1975 1980 1985

Number of Observations Read 447021

Number of Observations Used 447021

Dimensions

G-side Cov. Parameters 2

R-side Cov. Parameters 1

Columns in X 3

Columns in Z 21

Subjects (Blocks in V) 1

Max Obs per Subject 447021

Optimization Information

Optimization Technique Newton-Raphson with Ridging

Parameters in Optimization 2

Lower Boundaries 2

Upper Boundaries 0

Fixed Effects Profiled

Residual Variance Profiled

Starting From Data

Iteration History

Objective Max

Iteration Restarts Subiterations Function Change Gradient

0 0 2 216613.61122 0.09720382 0.059493

1 0 2 190505.17152 0.06191589 0.076546

2 0 1 190890.87879 0.00077464 0.000958

3 0 1 190890.9405 0.00000086 0.000017

4 0 0 190890.94117 0.00000000 0.000017

5 0 2 219914.01043 0.07439851 0.289449

6 0 2 189277.96071 0.07705141 0.173193

7 0 1 190887.97862 0.00346238 0.012421

8 0 1 190890.93974 0.00001771 2.324E-7

9 0 1 190890.94119 0.00000003 8.211E-7

10 0 0 190890.94117 0.00000001 4.33E-7

Convergence criterion (PCONV=1.11022E-8) satisfied.

Fit Statistics

-2 Res Log Pseudo-Likelihood 190890.9

Generalized Chi-Square 40101.08

Gener. Chi-Square / DF 0.09

Covariance Parameter Estimates

Standard

Cov Parm Estimate Error

co5 0.001237 0.000502

year 0.003539 0.002289

Residual 0.08971 0.000190

Solutions for Fixed Effects

Standard

Effect Estimate Error DF t Value Pr > |t|

Intercept 0.7630 0.02596 4 29.39 <.0001

ag5 -0.01842 0.002101 447E3 -8.77 <.0001

ag52 -0.00194 0.000088 447E3 -22.04 <.0001

Type III Tests of Fixed Effects

Num Den

Effect DF DF F Value Pr > F

ag5 1 447E3 76.86 <.0001

ag52 1 447E3 485.97 <.0001

Solution for Random Effects

reference Std Err

Effect year co5 Estimate Pred DF t Value Pr > |t|

co5 1915 -0.06919 0.01874 447E3 -3.69 0.0002

co5 1920 -0.01670 0.01591 447E3 -1.05 0.2939

co5 1925 -0.01789 0.01416 447E3 -1.26 0.2065

co5 1930 -0.00432 0.01262 447E3 -0.34 0.7324

co5 1935 0.000552 0.01128 447E3 0.05 0.9610

co5 1940 0.01859 0.01021 447E3 1.82 0.0685

co5 1945 0.04297 0.009505 447E3 4.52 <.0001

co5 1950 0.05873 0.009236 447E3 6.36 <.0001

co5 1955 0.04089 0.009439 447E3 4.33 <.0001

co5 1960 0.01730 0.01008 447E3 1.72 0.0862

co5 1965 0.009171 0.01110 447E3 0.83 0.4086

co5 1970 0.002871 0.01241 447E3 0.23 0.8170

co5 1975 -0.00982 0.01391 447E3 -0.71 0.4803

co5 1980 -0.02965 0.01558 447E3 -1.90 0.0570

co5 1985 -0.04351 0.01740 447E3 -2.50 0.0124

year 1985 -0.08876 0.02492 447E3 -3.56 0.0004

year 1990 -0.04037 0.02451 447E3 -1.65 0.0996

year 1995 -0.00430 0.02433 447E3 -0.18 0.8597

year 2000 0.01239 0.02433 447E3 0.51 0.6105

year 2005 0.04312 0.02451 447E3 1.76 0.0785

year 2010 0.07791 0.02486 447E3 3.13 0.0017

Note that the effect of “ag5” (standardized age in five-year brackets) is -0.01842 and the effect of “ag52” (standardized age squared in five year brackets) is -0.00194, meaning that age and age squared again have a negative effect on education credentials, in effect specifying that people lose their educational credentials when they age. However, while this effect on education must be incorrect, as we know that one cannot lose education with age, the age effect that previous papers have shown using APC-IE and HAPC are likely also mistaken, but interpreted to show substantive results. We have replicated the result with different countries and databases and conclude that the APCI-IE and HAPC give faulty age, and consequently, period and cohort effects.

Literature

Reither, Eric N., Robert M. Hauser, and Yang Yang. 2009. "Do birth cohorts matter? Age-period-cohort analyses of the obesity epidemic in the United States." *Social Science & Medicine* 69(10):1439-1448.

Yang, Yang, Sam Schulhofer Wohl, Wenjiang J. Fu, and Kenneth C. Land. 2008. "The Intrinsic Estimator for Age-Period-Cohort Analysis: What It Is and How to Use It." *American Journal of Sociology* 113(6):1697-1736.