**Lab session II – Wednesday – Model specification**

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

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

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

\*\*\*\* Chapter 10 -- Model specification \*\*\*\*

\*\*\*\* Stata do file to reproduce examples \*\*\*\*

\*\*\*\* in Chapter 10 of \*\*\*\*

\*\*\*\* Snijders, Tom A.B., and Bosker, Roel J. \*\*\*\*

\*\*\*\* Multilevel Analysis: \*\*\*\*

\*\*\*\* An Introduction to Basic and Advanced Multilevel Modeling,\*\*\*\*

\*\*\*\* second edition \*\*\*\*

\*\*\*\* London etc.: Sage Publishers, 2012 \*\*\*\*

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

\*\*\*\* Contributed by Jon Fahlander \*\*\*\*

\*\*\*\* jon.fahlander@nuffield.ox.ac.uk \*\*\*\*

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

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

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

\*Example 10.1 - level one heteroskedasticity

clear all

set more off

cd "J:\Multilevel TA"

\*infile schoolnr pupilNR\_new langpost ses IQ\_verb sex Minority denomina sch\_ses sch\_iqv sch\_min using "/Users/Jon/Documents/Snijders/SecondEditionExamples/DataSnijdersSecondEdition/mlbook2\_r.txt" in 2/3759, clear

\*infile schoolnr pupilNR\_new langpost ses IQ\_verb sex Minority denomina sch\_ses sch\_iqv sch\_min using mlbook2\_r.txt

use mlbook2\_r.dta, clear

gen IQXSES = IQ\_verb\*ses

bysort schoolnr: gen groupsize = \_N

sort groupsize

egen schoolRelSize = group(school)

drop if groupsize<10

egen newSchoolnr = group(schoolnr)

sum newSchoolnr

display r(max) // OBS! For some reason the data contains 180 groups, not 133 groups, with 10 students or more.

sum newSchoolnr

local nrgroups=r(max)

matrix M= J(`nrgroups',1,.)

matrix D= J(`nrgroups',1,.)

matrix list M

matrix list D

foreach i of num 1/`nrgroups' {

reg langpost IQ\_verb ses IQXSES sex if newSchoolnr == `i'

local dfXlogRSS =e(df\_r)\*log(e(rss))

local df=e(df\_r)

matrix M[`i',1]=`dfXlogRSS'

matrix D[`i',1]=`df'

}

mat list M

mat list D

\*sum all the columns

mat rowvectorof1s = J(rowsof(M),1,1)

mat list rowvectorof1s

mat sumM=rowvectorof1s'\*M

mat sumD=rowvectorof1s'\*D

\*The nominator and denominator of equation 10.3

mat list sumM

mat list sumD

\*The result, "ls\_tot" in equation 10.3

display sumM[1,1]/sumD[1,1]

\*Equation 10.4

sum newSchoolnr, meanonly

local nrgroups=r(max)

display `nrgroups'

matrix djs= J(`nrgroups',1,.)

foreach i of num 1/`nrgroups' {

reg langpost IQ\_verb ses IQXSES sex if newSchoolnr == `i'

local dj = (e(df\_r)/2)^(1/2)\*(log(e(rss))-sumM[1,1]/sumD[1,1])

matrix djs[`i',1]=`dj'

}

matrix list djs

\*Equation 10.5

sum newSchoolnr, meanonly

local nrgroups=r(max)

display `nrgroups'

matrix dj2s= J(`nrgroups',1,.)

foreach i of num 1/`nrgroups' {

matrix dj2s[`i',1]=djs[`i',1]^2

}

matrix list dj2s

matrix list djs

mat sumdj2s =rowvectorof1s'\*dj2s

matrix list sumdj2s

local H=sumdj2s[1,1]

display `H'

\*chisquare test, with a teststatistic of 490 and df 180-1

local x=chi2(`nrgroups'-1,`H')

local x=chi2(180-1, 489.88)

display 1-`x'

\*test in a chisquare dist with df 132 and a teststatistic of 155

local x=chi2(133-1,155)

display 1-`x' //Higly significant

\*To visually inspect the heterogeneity group level standard deviations one could make a graph of the distribution of within group residual sd

qui: sum newSchoolnr

local nrgroups=r(max)

matrix sds= J(`nrgroups',1,.)

matrix list sds

foreach i of num 1/`nrgroups' {

reg langpost IQ\_verb ses IQXSES sex if newSchoolnr == `i'

local groupSD = e(rmse)

matrix sds[`i',1]=`groupSD'

}

svmat sds

hist sds, norm

\*Example 10.2 -- Inspection of level one residuals

clear

\*infile schoolnr pupilNR\_new langpost ses IQ\_verb sex Minority denomina sch\_ses sch\_iqv sch\_min using "/Users/Jon/Documents/Snijders/SecondEditionExamples/DataSnijdersSecondEdition/mlbook2\_r.txt" in 2/3759, clear

use mlbook2\_r.dta, clear

set more off

gen IQXSES = IQ\_verb\*ses

bysort schoolnr: gen groupsize = \_N

sort groupsize

egen schoolRelSize = group(school)

drop if groupsize<10

egen newSchoolnr = group(schoolnr)

sum newSchoolnr

display r(max) // OBS! For some reason the data contains 180 groups with 10 observations or more, not 133 groups, with 10 observations or more.

\*gen IQ\_verb2=round(IQ\_verb,0.5) // Round IQ to nearest 0.5

\*Calculate and store the residuals from the within group regressions

gen WithinGroupRes=.

su newSchoolnr, meanonly

local nrgroups=r(max)

matrix sds= J(`nrgroups',1,.)

matrix list sds

foreach i of num 1/`nrgroups' {

reg langpost IQ\_verb ses IQXSES sex if newSchoolnr == `i'

predict residuals`i', res

replace WithinGroupRes=residuals`i' if newSchoolnr == `i'

drop residuals`i'

}

\*Drop observations if less than 12 in each IQ\_verb category

bysort IQ\_verb: gen obsInEachIQcat = \_N

\*drop if obsInEachIQcat <12

egen meanRes = mean(WithinGroupRes) if obsInEachIQcat >11, by(IQ\_verb)

egen sdRes = sd(WithinGroupRes) if obsInEachIQcat >11, by(IQ\_verb)

gen Upper2sd = meanRes + 2\*sdRes if obsInEachIQcat >11

gen Lower2sd = meanRes - 2\*sdRes if obsInEachIQcat >11

\*Create a graph of the mean residuals across IQ (figure 10.1 A)

scatter meanRes IQ\_verb if obsInEachIQcat >11 || rcap Upper2sd Lower2sd IQ\_verb if obsInEachIQcat >11

\*Create a graph of the mean residuals across ses (figure 10.1 B)

tab ses

gen ses\_2\_cat=round(ses,2)

tab ses\_2\_cat

egen meanRes\_ses = mean(WithinGroupRes) , by(ses\_2\_cat)

egen sdRes\_ses = sd(WithinGroupRes) , by(ses\_2\_cat)

gen Upper2sd\_ses = meanRes\_ses + 2\*sdRes\_ses

gen Lower2sd\_ses = meanRes\_ses - 2\*sdRes\_ses

\*Create a graph of the mean residuals across ses\_2\_cat

scatter meanRes\_ses ses\_2\_cat || rcap Upper2sd\_ses Lower2sd\_ses ses\_2\_cat

\*Reestimate the model with non-linear effects of IQ

gen IQ2=IQ\_verb^2

gen IQ3=IQ\_verb^3

replace IQ2 =0 if IQ\_verb<0

replace IQ3 =0 if IQ\_verb<0

gen WithinGroupRes2=.

su newSchoolnr, meanonly

local nrgroups=r(max)

matrix sds= J(`nrgroups',1,.)

matrix list sds

foreach i of num 1/`nrgroups' {

reg langpost IQ\_verb IQ2 IQ2 ses IQXSES sex if newSchoolnr == `i'

predict residuals`i', res

replace WithinGroupRes2=residuals`i' if newSchoolnr == `i'

drop residuals`i'

}

\* Calculate the mean residuals in each cat (figure 10.2 A and B)

\*IQ

egen meanResIQ2 = mean(WithinGroupRes2) if obsInEachIQcat >11, by(IQ\_verb)

egen sdResIQ2 = sd(WithinGroupRes2) if obsInEachIQcat >11, by(IQ\_verb)

gen Upper2sdIQ2 = meanResIQ2 + 2\*sdResIQ2 if obsInEachIQcat >11

gen Lower2sdIQ2 = meanResIQ2 - 2\*sdResIQ2 if obsInEachIQcat >11

\*The new IQ graph (figure 10.2 A)

scatter meanResIQ2 IQ\_verb if obsInEachIQcat >11 || rcap Upper2sdIQ2 Lower2sdIQ2 IQ\_verb if obsInEachIQcat >11

\*ses

egen meanRes\_ses2 = mean(WithinGroupRes2) , by(ses\_2\_cat)

egen sdRes\_ses2 = sd(WithinGroupRes2) , by(ses\_2\_cat)

gen Upper2sd\_ses2 = meanRes\_ses2 + 2\*sdRes\_ses2

gen Lower2sd\_ses2 = meanRes\_ses2 - 2\*sdRes\_ses2

\*The new ses-graph (figure 10.2 B)

scatter meanRes\_ses2 ses\_2\_cat || rcap Upper2sd\_ses2 Lower2sd\_ses2 ses\_2\_cat

\*Example 10.3

clear

\*infile schoolnr pupilNR\_new langpost ses IQ\_verb sex Minority denomina sch\_ses sch\_iqv sch\_min using "/Users/Jon/Documents/Snijders/SecondEditionExamples/DataSnijdersSecondEdition/mlbook2\_r.txt" in 2/3759, clear

use mlbook2\_r.dta, clear

set more off

\*Regenerate the auxilliary variables

gen IQXSES = IQ\_verb\*ses

egen gmeanIQverb = mean(IQ\_verb) , by(schoolnr)

egen gmeanSES = mean(ses), by(schoolnr)

gen IQMeanXsesMean=gmeanSES \* gmeanIQverb

gen IQ2=IQ\_verb^2

gen IQ3=IQ\_verb^3

replace IQ2 =0 if IQ\_verb<0

replace IQ3 =0 if IQ\_verb<0

\*The regression model of interest

xtmixed langpost IQ\_verb IQ2 IQ3 ses IQXSES sex gmeanIQverb gmeanSES IQMeanXsesMean || schoolnr: IQ\_verb

predict lev2res\*, reffects level(schoolnr)

scatter lev2res2 gmeanIQverb

\*The qqplot (figure 10.3)

egen stdResIntercept=sd(lev2res2)

gen stizedIntercept= lev2res2/stdResIntercept

egen stdResSlope=sd(lev2res1)

gen stizedSlope= lev2res1/stdResSlope

qnorm stizedSlope, name(Slope)

qnorm stizedIntercept, name(Intercept)

\*The IQ spline

lowess lev2res2 gmeanIQverb

lowess lev2res1 gmeanIQverb

\*The ses spline

lowess lev2res2 ses

lowess lev2res1 ses

\*Example 10.4 Influence of level-two units

\*At the moment there is no built in method to calculate the cooks distance of the level two units.

\*There is however a beta version of a user package called mlt (multilevel tools)

\*The package is found here

\*http://www.alexanderwschmidt.de/multileveltools.html

\*Once installed you should able to get the influence statistics discussed in example 10.4 by typing

mltcooksd, random keepvar(cook)