version 7
cd "c:\statawrk\Women\"
capture log close
log using Oct28.log, replace

**************************************************************;
* Name:  reestimate.do
* Purpose: Our version of the CDGLM modelling.
* Version:
* Modified: October 2003
* Author: T Breusch & E Gray
* Infile: NLC Wave 1 main file c:\statawrk\nlc\wavel\d1015.sta
* *
* Outfile: logfile only
* Stata ver: 7
**************************************************************;

#delimit ;
clear;
set more off;
set memory 12m;

*FIRST PASS - Replicate CDGLM Statistics ********************
* Attempts to match Matthew Gray’s code, where possible, are
* shown by notes like this *### Page 2, line 21 ###*

use "c:\statawrk\nlc\Wavel\d1015.dta", clear;

* Drop males and self employed;
drop if q22==1; drop if q81==3;

* Codings for Table 1;
* Age of respondent in groups;
gen agegpr=.;
replace agegpr=1 if ager>=18&ager<=24;
replace agegpr=2 if ager>=25&ager<=34;
replace agegpr=3 if ager>=35&ager<=44;
replace agegpr=4 if ager>=45&ager<=55;

label define agegpr
1 "18-24"
2 "25-34"
3 "35-44"
4 "45-55";

label values agegpr agegpr;
label var agegpr "age group of respondent";

* Employed apparently means actually worked last week (q77==1), or on
* some leave (q77==2 or 3) and recorded positive hours.
* ### Page 2, line 21 ###;
gen employed=(q77==1);
replace employed=1 if (q77==2&q77==3)&q110>0;

* Person id=32 seems to be declared not employed;
replace employed=0 if id==32;

* Actual hours worked for those who are employed;
gen hours=0;
replace hours=q110 if employed==1&q110>0;

******************************
* Children in the household;
******************************
* Count the children living at home (q11aX=1)
* Create ch15 = n for n children <15
* ch03 = n for n children <3 yrs
* ch03_n = 1 if n children <3 yrs, n=1,2 or more
* ch15_n = 1 if n children 3-15 yrs, n=1,2,3 or more;
gen byte chnn=0;
gen byte ch15=0;
gen byte ch03=0;

* This is what replicates their numbers! Note there is no check for
* living at home, and the age ranges are <5 and 5-14;
for num 1/8, noh:
    replace chnn=chnn+1 if agechX>=0&agechX~=. 
    replace ch15=ch15+1 if agechX>=0&agechX<15 
    replace ch03=ch03+1 if agechX>=0&agechX<5;

gen ch03_1=(ch03==1);
gen ch03_2=(ch03==2);
gen ch15_1=(ch15-ch03==1);
gen ch15_2=(ch15-ch03==2);
gen ch15_3=(ch15-ch03>=3);

*********************************************
* Examine work history to construct experience variable;
*********************************************
* Count years the person has been working ft and pt;
gen byte wkft=0;
gen byte wkpt=0;
for num 55/96, noh: qui replace wkft=wkft+1 if q14aX==1;
for num 55/96, noh: qui replace wkpt=wkpt+1 if q14aX==2;

label var wkft "count yrs work ft";
lable var wkpt "count yrs work pt";
gen exper=wkft+wkpt*20/35;

*********************************************
* Code some other variables (names mostly taken from CDGLM);
*********************************************
* The variable "Number of children" in Table 2 seems to be from
* Q154 "How many children have you ever had";
gen chnx=q154 if q154<=3;
replace chnx=3 if q154>3;
* This measure of tenure gets close, but what's it mean?;
    gen tenure=q100 if q100>=0;
replace tenure=. if q100<0&employed==1;
replace tenure=0 if q100<0&employed==0;
gen tenure2=tenure*tenure;
gen exper2=exper*exper;
gen age24=(ager<=24);
gen age34=(ager>=25&ager<=34);
gen mar=(q20==3|q20==4);
* This is apparently how CDGLM do it;
gen degree=(q57_rec==1|q57_rec==2|q57_rec==3);
gen trade=(q57_rec==4|q57_rec==5|q57_rec==6|q57_rec==7);
gen yr12=(q52==1&degree==0&trade==0);
gen nesb=(q29~=1&q29~=5&q29~=14&q29~=24&q29~=25&q29~=38&q29~=43&q29~=58&q29~=59&q29~=61);
### They call Singapore non-NESB ###;
replace nesb=0 if q29==46;
gen everch=(ceb>0);
* These are wrong-way round, to replicate CDGLM;
gen breadw=(q267a1==1|q267a1==2);
gen contrib=(q234a3==4|q234a3==5);

******************************************************;
* Net wage (on assumption wage is the only income);
******************************************************;
    gen wnet=wage;
replace wnet=wage-0.2*(wage-5400) if wage>5400&wage<=20700;
replace wnet=wage-3060-0.34*(wage-20700) if wage>20700&wage<=38000;
replace wnet=wage-8942-0.43*(wage-38000) if wage>38000&wage<=50000;
replace wnet=wage-14102-0.47*(wage-50000) if wage>50000;
******************************************************;
* Replicate Table 1 & Table 2?;
******************************************************;
by agegr, s: summ employed wnet;
by agegr, s: summ hours wnet if employed==1;
by chnx, s: summ ager employed wnet;

******************************************************;
* Replicate Table A1?;
******************************************************;
summ exper exper2 tenure tenure2 age24 age34 mar degree trade
yr12 nesb everch ch03_1 ch03_2 ch15_1 ch15_2 ch15_3
pinc contrib breadw;

**********************************************************************;
####### SECOND PASS - Re-estimate on CDGLM Definitions #######;
**********************************************************************;
#delimit ;
use "c:\statawrk\nlc\Wave1\d1015.dta", clear;

* Drop males and self employed;
drop if q22==1; drop if q81==3;

***************************************************************;
* Children in the household;
***************************************************************;
* Count the children living at home (q11aX=1)
* Create ch15 = n for n children <=15
* ch03 = n for n children <3 yrs
* ch03_n = 1 if n children <3 yrs, n=1,2 or more
* ch15_n = 1 if n children 3-15 yrs, n=1,2,3 or more;
gen byte ch15=0;
gen byte ch03=0;

* This follows CDGLM definitions for numbers of children in age groups;
for num 1/8, noh:
   replace ch15=ch15+1 if agechX>=0&agechX<=15&q11aX==1 \
   replace ch03=ch03+1 if agechX>=0&agechX<3&q11aX==1;
gen ch03_1=(ch03==1);
gen ch03_2=(ch03>2);
gen ch15_1=(ch15-ch03==1);
gen ch15_2=(ch15-ch03==2);
gen ch15_3=(ch15-ch03>=3);

***************************************************************;
* Fix some variables. Create some new variables;
***************************************************************;
* AGE OF RESPONDENT One person has age missing (assume birthday
* before interview date of 03-NOV-96);
   replace ager=1997-1950 if id==867;
   replace age5gp=7 if id==867;

* EMPLOYED means actually worked last week. Drop the bad cases
* from sample later;
gen employed=.;

* Employed if worked last week;
   replace employed=1 if q77==1;

* Employed if on maternity leave and positive hours worked;
   replace employed=1 if q77==2&q110>0;
   replace employed=0 if q77==2&q110==0;

* Employed if on other leave (typically short-term);
   replace employed=1 if q77==3;

replace employed=0 if q77==4;

* TENURE based on q100 unless q77=4;
gen tenure=q100 if q100>0;
replace tenure=0 if q77==4;
replace tenure=0 if q77==2&employed==0;

* HOURS based on q110;
gen hours=q110 if employed==1;
replace hours=0 if employed==0;

* WAGE income, based on wage;
gen winc=wage;
replace winc=0 if employed==0;

* EXPERIENCE from years the person has been working ft and pt.
* Examine work history to construct experience variable;
gen byte wkft=0;
gen byte wkpt=0;
for num 55/96, noh: qui replace wkft=wkft+1 if q14aX==1;
for num 55/96, noh: qui replace wkpt=wkpt+1 if q14aX==2;
label var wkft "count yrs work ft";
label var wkpt "count yrs work pt";
gen exper=wkft+wkpt*20/35;

******************************************************************************;
* Code some other variables;******************************************************************************;
gen ten200=tenure*tenure/100;
gen exp200=exper*exper/100;
gen age24=(ager<=24);
gen age34=(ager>=25&ager<=34);
gen mar=(q20==3|q20==4);
gen degree=(highed==1|highed==2);
gen trade=(highed==3);
gen yr12=(highed==4);
gen nohs=(degree==0&trade==0&yr12==0);
gen nesb=(q29~~=1&q29~~=5&q29~~=14&q29~~=24&q29~~=25&q29~~=38&q29~~=43&q29~~=58&q29~~=59&q29~~=61);
gen everch=(ceb>0);
gen contrib=(q267a1==1|q267a1==2);
gen breadw=(q234a3==1|q234a3==2);
replace pinc=pinc/1000;
gen age200=ager*ager/100;

******************************************************************************;
* Net Income;******************************************************************************;
* Net wage (on assumption wage is the only income);
gen wnet=winc;
replace wnet=winc-0.2*(winc-5400) if winc>5400&winc<=20700;
replace wnet=winc-3060-0.34*(winc-20700) if winc>20700&winc<=38000;
replace wnet=winc-8942-0.43*(winc-38000) if winc>38000&winc<=50000;
replace wnet=winc-14102-0.47*(winc-50000) if winc>50000;

gen lwnet=100*log(wnet);

************************************************************;
* Table 1 - Re-estimation of CDGLM summary statistics;
************************************************************;
summ exper exp200 tenure ten200 age24 age34 mar degree trade
yr12 nesb everch ch03_1 ch03_2 ch15_1 ch15_2 ch15_3
pinc contrib breadw ager age200 employed wnet;

******************************************************;
* Drop the bad cases; 
******************************************************;
* Drop from sample if tenure is missing (Q100<0); 
* Drop if the implied hourly rate is implausibly small, 
* (<$1, but <$0.50 or <$2 will have exactly the same effect; 
* Drop if employed==1&(wage/(q110*52)<1);

drop if winc==.;

*************************************************************;
* Table 2 - Summary statistics for women age 20-55; 
*************************************************************;
summ exper exp200 tenure ten200 age24 age34 mar degree trade
yr12 nesb everch ch03_1 ch03_2 ch15_1 ch15_2 ch15_3
pinc contrib breadw ager age200 wnet if ager>=20&employed==0;
summ exper exp200 tenure ten200 age24 age34 mar degree trade
yr12 nesb everch ch03_1 ch03_2 ch15_1 ch15_2 ch15_3
pinc contrib breadw ager age200 wnet if ager>=20&employed==1;
summ exper exp200 tenure ten200 age24 age34 mar degree trade
yr12 nesb everch ch03_1 ch03_2 ch15_1 ch15_2 ch15_3
pinc contrib breadw ager age200 employed wnet if ager>=20;

*************************************************************;
* Table 3 - Estimation results for women age 20-55; 
*************************************************************;
set matsize 150;
heckman lwnet exper exp200 tenure ten200 ager age200 mar degree trade
yr12 nesb everch ch03_1 ch03_2 ch15_1 ch15_2 ch15_3 pinc if ager>=20,
select (employed=exper exp200 ager mar degree trade yr12 nesb everch
ch03_1 ch03_2 ch15_1 ch15_2 ch15_3 pinc contrib breadw) robust;

*************************************************************;
* Some calculations for initializing the scenarios;
summ employed if age5gp==2&degree==1;
summ employed if age5gp==2&yr12==1;
summ employed if age5gp==2&nohs==1;

log close;