Gillian Paull
Institute for Fiscal Studies
June 2003
Revised: November 2010

BHPS Employment Histories: User Guide

The first section of this guide provides an introductory description of the construction of the data. The second section lists the data sets and their dimensions, the third provides a table of variables and the final section presents the STATA do-files used to construct the data from the original BHPS data.

1. Description

The description contained in this section is a brief overview of the construction of the datasets. A full discussion of the details can be found in "Biases in the Reporting of Labour Market Dynamics", Institute for Fiscal Studies Working Paper WP02/10 (available at http://www.ifs.org.uk/workingpapers/wp0210.pdf). Consultation of this document is recommended before use of the employment history data.

The data set consists of eight data files derived from the first 11 waves of the British Household Panel Survey (BHPS). Each file contains the same sample and variables, but the variables have been constructed using a different method in each file. The unit of observation is each interview provided by each individual during the first 11 waves of the BHPS[1]. Each record contains a series of labour forces spells since leaving full-time education until the time of the interview. The spells are defined in terms of spell state, start date and end date. Self-employment spells are included as separate employment spells. Spells are in chronological order and all dates are measured in months, beginning with month 1 as January 1900. For example, September 1991 is month 1101. Summary measures of the number of months in employment since leaving full-time education and employer tenure in months have also been derived from the series of spells. Missing or inconsistent data are indicated by the variables denoted *prob. The individual identifier (pid) and wave identifier (intwave) are included in all datasets to allow the data to be merged across the datasets and with other data from other BHPS sources.

Two types of data source and four construction methods define the eight data sets.

The two types of data sources are:
(1): "Main Activity". This uses information from the individual's self-defined main activity and uses the BHPS variables:

- variables jbstat, nemst, cjsbgm/y in data file INDRESP

[1] In the analysis in the working paper, observations from the ECHP sample for wave 7 onwards and from Wales and Scotland from wave 9 onwards were dropped in order to maintain a representative sample. These observations have been included in the employment history files in order to allow the researcher discretion on whether to use them. Observations from Northern Ireland for wave 11 are also included.

    - priority is given to jbstat in wave A and to nemst in subsequent waves as these correspond to the starting date reported
- variables jhstat, jhsemp, jhbgd/m in data file JOBHIST
- variables bleshst, bleshem/y in data file BLIFEMST
- variables cljsemp, cljbgm/y, cljlftm/y from data file CLIFEJOB to add employer-to-employer moves prior to September 1990

(2): "All-work". This uses information collected on any employment, regardless of whether it is the main activity. It uses the BHPS variables:

- variables jbsemp, jbbgm/y, jsbgm/y from data file INDRESP
- variables jhstat, jhsemp, jhbgd/m in data file JOBHIST
    - strictly speaking, this is a main activity variable because jobs held between interviews which are not regarded as the main activity are not picked up in the survey
- variables cljsemp, cljbgm/y, cljlftm/y from data file CLIFEJOB

The four different methods for the data construction derive from the fact that the BHPS collects information at each interview on labour force spells extending prior to the previous interview. This can generate inconsistencies between interviews in the spells reported during this "overlap" period. The four methods of construction are based on which source of information is used for this overlap period:

Method A: Traditional panel: ignores any information prior to previous interview
Method B: Latest interview rules: only uses information from previous interviews to complete gaps
Method C: Reconciled: case-by-case correction of inconsistencies
Method D: Selected no problems. Experience and tenure variables only non-missing for cases where there were no inconsistencies or missing infromation in the original data required for constructing these variables. Spell data should be selected by prob*=0.

2. Data Files

The data files are named hist*# where * denotes the method used and # the data source.

Data file	Number of observations	Number of variables	Maximum number of spells per observation	Size
histA1	125,116	254	48	128 mk
histB1	125,116	209	39	105 mk
histC1	125,116	244	46	123 mk
histD1	125,116	254	48	128 mk
histA2	125,116	219	41	110 mk
histB2	125,116	154	28	78 mk
histC2	125,116	209	39	105 mk
histD2	125,116	219	41	110 mk

3. Variable List

Missing values use the usual STATA "." unless otherwise specified.

Variable	Description
pid	personal identifier from BHPS
method	A. traditional panel B. latest interview rules C. reconciled D. selected no problems
datasource	1: main activity 2: all work
intwave	wave 1-11
intdate	interview date in months with January 1900=1
currstat	current state: 1 self-employed 2 employed 3 unemployed 4 out of labour force 7 full-time education 99 missing
tenure	employer tenure in months or number of months in self-employment spell
exper	employment experience in months
left	month first left full-time education (from bledendm & bledendy)
smin	start date for first spell (start of history)
total	total number of spells
wave# (#=spell)	first wave that spell appears in: 0-11
stat# (#=spell)	spell status: 1 self-employed 2 employed 3 unemployed 4 out of labour force 7 full-time education 99 missing
sdate# (#=spell)	month spell started
edate# (#=spell)	month spell ended (= month next spell started)
prob# (#=spell)	problems with spell (or combination) : 0: no problems 1: previous spell has missing stat 2: missing stat 3: next spell has missing stat 20: missing start date 200: missing end date 2000: start date > end date 10000: overlap with previous spell 30000: overlap with next spell 100000: gap with previous spell and (source 1 or wave>0) 300000: gap with next spell and (source 1 or wave>0)
tprob	problems in tenure calculation (or combination): 0: no problems 1: current work spell has a problem 10: current work spell is first spell and begins after wave 0 and after date left FT educ
eprob	problems in experience calculation (or combination): 0: no problems #: number of work spells with a problem plus number of spells with a missing state 100: no wave B employment history and date left full-time education missing or before first spell start date
sprob	start date for spells with no problems: 0: no problems intdate: last spell has a problem

4. STATA Do-Files Used to Construct Data

This final section aims to provide greater detail on the construction of the data sets for those familiar with STATA programming. It also provides the opportunity to extend the construction of the employment histories beyond wave 11 as further data becomes available.

The employment history data was constructed using three STATA do-files: bhpshist1.do, bhpshist2.do, bhpshist3.do. Each do-file consists of a number of sub-programs that are called at the end of the main file. The summaries on pages 4-7 list the steps taken in each program in each do-file. The remainder of the section reproduces the STATA do-files. The text for the STATA do-files can be cut and pasted into text editors and run in STATA. Each of the three do-files should be run sequentially. Directory references should obviously be amended appropriately. Employment histories can be constructed for additional waves of the BHPS using the editing instructions at the beginning of bhpshist1.do.

Summary of bhpshist1.do

ReadData
- reads the raw data from indresp, jobhist, blifemst and clifejob
- keeps only full interview observations (drops proxies)
- adds pid to jobhist data
- recodes all negative values to missing
- corrects some year dates in clifejob from 1900
- saves as bhpsw1 bhpsw2 blifemst clifejob
- saves educend with dates first left full-time eduation

ACorrect
- reorders jbstat and jhstat for wave A to match subsequent waves

WSpells
- combines and amends spells form indresp and jobhist
- creates interview date for indresp in months (Jan 1990 = 1)
- drops observation if idate missing (20 cases in wave A)
- creates starting dates, with seasons replaced as January (winter), April (Spring), July (Summer), October (Autumn) and missing months as July for years prior to 1990
- creates a work variable for indresp based on jbhas jboff and jbsemp
- drops spells which begin after the interview date for indresp
- the starting date for main activity in indresp is replaced as missing for non-students if some work is reported and cjsdate equals the work starting date. It appears that all work starting dates have been copied to cjsdate regardless of jbstat
- status and starting date for indresp selected to correspond with the break date for previous spell:
  - main activity for students and for non-students without any work
  - work for non-students with any work
- realigns status in jobhist and sets status = 1.5 for different job/same employer
- saves spells and tempadd

EmpMerge
- merges different job/same employer with subsequent employment spell
- saves wspells

BLife
- adjusts errors in bleshno so that spell numbers run consecutively

- adjusts status to match indresp
- creates end dates and start date for first spell in months
- saves blife

CLife
- adjusts status to match indresp
- creates start dates and end dates in months
- saves clife

Demogs
- creates a data set of demographic information
- combined with tempadd to save addinfo

Summary of bhpshist2.do

BAmend
- adds start dates as previous spell end dates
- converts status to 5 categories
- merges consecutive spells of the same state
- saves tempb

CAmend
- reshapes long
- save tempc

BCAmend
- identifies employer-to-employer moves in clife and whether they fall within employment spells in tempb
- selects individuals in both blifemst and clifejob *or* individuals in blifemst with no employment spell prior to September 1990
- splits employment spells in tempb by the matched employer moves
- saves checkbc and tempbc

WMain
- indresp spell status set to main activity and starting date set to missing for those with non-work main activity and reporting some work
- saves tempmain

WWork
- indresp spell status set to work and starting date set to missing for full-time students also reporting some work
- saves tempwork

WAmend
- adds end dates as subsequent spell start dates or interview date
- converts status to 5 categories, including remaining 1.5 in employment and using nemst if there is a conflict between jbstat and nemst
- merges consecutive spells of the same state unless the state is employment
- saves wavemain and wavework

BPanel
- combines tempb (as wave 0) with wavemain
- spells in wave 0 dropped if start after August 1990 (or August 1991 if not interviewed in wave A)
- saves bpanel

CPanel
- combines tempc (as wave 0) with wavework
- spells in wave 0 dropped if start after August 1990 (or August 1991 if not interviewed in wave A)

- saves cpanel

BCPanel

- combines tempbc (as wave 0) with wavemain
- spells in wave 0 dropped if start after August 1990 (or August 1991 if not interviewed in wave A)
- saves bcpanel

Sumamry of bhpshist3.do

In this final do-file, each method of construction calls different sub-programs as follows:

Method:	A	B	C	D	
Calls sub-program:					
DateFixEM/DateFixW	-	-	yes	-	
CrosWave	yes	-	yes	yes	
SeamsA	yes	-	-	-	
ReshapeB	-	yes	-	-	
Divide	yes	yes	yes	yes	
AmendC		-	-	yes	-
SelectB	-	yes	-	-	
Problems / ExperTen	yes	yes	yes	yes	

DateFixEM/DateFixW

- amends imputed dates

CrosWave/SubCW

- drops spells which end before prior interview
- merges spells matched across waves

SeamsA/SubSA

- creates false break for cross-wave seams

ReshapeB

- reshapes data

Divide

- divide into interviews, keeping spells from that and previous interviews

AmendC / SubAmend

- missing start dates for spell1 replaced with date left education if within 12 months of edate for spell 1
- creates a break date for spells with missing end and start dates (noting that a missing end date for spell i must correspond to a missing start date for a spell j within the first wave that spell j appears in)
- extends spell if there is a gap with the next spell and the spells are in the same or consecutive waves
- merges spells with missing state which overlap
- merges spells with missing state and missing date but close other date
- merges spells with missing end date which overlap by start dates and have same state
- merges spells with missing end date which have same state~=2
- merges spells with same end dates and stat but missing start date
- merges spells with end dates within 12 months and same stat but missing start date
- creates a break for spells with end dates within 12 months and different stat but missing start date
- drops spell 1 if stat1 is full-time education
- drops second spell of spells which overlap and have same end date

- creates a break date for overlapping spells of the end date of the first spell
- initial spells dropped if spells of full-time education

SelectB/SubSB

- selects last spell and set cut=sdate
- drop previous spell if sdate>cut
- merge previous spell if match state=2/3 or match state=1 and edate>cut
- replace cut=sdate if previous spell not dropped

Problems

- identify spells with problems

ExperTen

- adds experience and tenure variables

```
*******************************************************************************
*******************************************************************************
*******************************************************************************

/* bhpshist1: Reads the raw BHPS data and first step in creating the basic data */
/*
To add subsequent waves to the histories requires this file to be edited in the
following way:
- in the program "ReadData", the local macros and loops around waves b-k and a-k need
to be extended for each additional wave
- in the program "WSpells", the local macros and 2 loops around b-k need to be
extended for each additional wave
The remaining programs automatically adjust for any increases in the number of spells
*/

#delimit ;
set more 1 ;
cd c:\gillian\disk10ext ;
log using bhpshist1.log, replace ;


************************* ReadData ******************************** ;
/* Reads in the data */

capture program drop ReadData ;
program define ReadData ;

local let1 "a" ; local let2 "b" ; local let3 "c" ; local let4 "d" ;
local let5 "e" ; local let6 "f" ; local let7 "g" ; local let8 "h" ;
local let9 "i" ; local let10 "j" ; local let11 "k" ;

/* reading indresp */
local i=0 ;
/* loop around wave a */
while `i' ==0 { ;
       local i = `i' +1 ;
       use pid `let`i''ivfio `let`i''hid `let`i''pno `let`i''doid `let`i''doim
       `let`i''jbstat `let`i''jbhas `let`i''jboff `let`i''jboffy `let`i''jbsemp
       `let`i''cjsbgd `let`i''cjsbgm `let`i''cjsbgy `let`i''jbbgd `let`i''jbbgm
       `let`i''jbbgy `let`i''jsbgd `let`i''jsbgm `let`i''jsbgy `let`i''sex
       `let`i''dobm `let`i''doby `let`i''feend `let`i''scend `let`i''age
       `let`i''qfedhi using m:\bhps11\data\\`let`i''indresp ;
       keep if `let`i''ivfio==1 ; drop `let`i''ivfio ;
       renpfix `let`i'' ; gen wave=`i' ;
       quietly mvdecode _all, mv(-9) ; quietly mvdecode _all, mv(-8) ;
       quietly mvdecode _all, mv(-7) ; quietly mvdecode _all, mv(-6) ;
       quietly mvdecode _all, mv(-5) ; quietly mvdecode _all, mv(-4) ;
       quietly mvdecode _all, mv(-3) ; quietly mvdecode _all, mv(-2) ;
       quietly mvdecode _all, mv(-1) ;
       save bhps`let`i''1.dta, replace ; drop _all ;
       } ;
local i=1 ;
/* loop around waves b to k */
while `i' <=10 { ;
       local i = `i' +1 ;
       use pid `let`i''ivfio `let`i''hid `let`i''pno `let`i''doid `let`i''doim
       `let`i''doiy `let`i''jbstat `let`i''nemst `let`i''jbhas `let`i''jboff
       `let`i''jboffy `let`i''jbsemp `let`i''cjsbgd `let`i''cjsbgm `let`i''cjsbgy
       `let`i''jbbgd `let`i''jbbgm `let`i''jbbgy `let`i''jsbgd `let`i''jsbgm
       `let`i''jsbgy `let`i''sex `let`i''dobm `let`i''doby `let`i''feend
       `let`i''scend `let`i''age `let`i''qfedhi
       using m:\bhps11\data\\`let`i''indresp ;
       keep if `let`i''ivfio==1 ; drop `let`i''ivfio ;
       renpfix `let`i'' ; gen wave=`i' ;
       quietly mvdecode _all, mv(-9) ; quietly mvdecode _all, mv(-8) ;
       quietly mvdecode _all, mv(-7) ; quietly mvdecode _all, mv(-6) ;
       quietly mvdecode _all, mv(-5) ; quietly mvdecode _all, mv(-4) ;
       quietly mvdecode _all, mv(-3) ; quietly mvdecode _all, mv(-2) ;
       quietly mvdecode _all, mv(-1) ;
       save bhps`let`i''1.dta, replace ; drop _all ;
       } ;

/* reading jobhist */
local i=0 ;
/* loop around waves a to k */
while `i'<=10 { ;
```

```
        local i=`i'+1 ;
        use `let`i''hid `let`i''pno `let`i''jspno `let`i''jhstat `let`i''jhsemp
        `let`i''jhbgd `let`i''jhbgm `let`i''jhbgy
         using m:\bhps11\data\\`let`i''jobhist ;
        renpfix `let`i'' ; gen wave=`i' ;
        quietly mvdecode _all, mv(-9) ; quietly mvdecode _all, mv(-8) ;
        quietly mvdecode _all, mv(-7) ; quietly mvdecode _all, mv(-6) ;
        quietly mvdecode _all, mv(-5) ; quietly mvdecode _all, mv(-4) ;
        quietly mvdecode _all, mv(-3) ; quietly mvdecode _all, mv(-2) ;
        quietly mvdecode _all, mv(-1) ;
        sort wave hid pno ; save bhps`let`i''2.dta, replace ; drop _all ;
         use bhps`let`i''1 ; keep wave pid hid pno ; sort wave hid pno ;
        merge wave hid pno using bhps`let`i''2 ; drop if pid==. ;
        save bhps`let`i''2 , replace ;
        drop _all ;
        } ;

/* reading blifemst */
use bhid bpno bleshno bleshst bleshsm bleshsy bleshem bleshey bleshne using
m:\bhps11\data\blifemst ;
quietly mvdecode _all, mv(-9) ; quietly mvdecode _all, mv(-8) ;
quietly mvdecode _all, mv(-7) ; quietly mvdecode _all, mv(-6) ;
quietly mvdecode _all, mv(-5) ; quietly mvdecode _all, mv(-4) ;
quietly mvdecode _all, mv(-3) ; quietly mvdecode _all, mv(-2) ;
quietly mvdecode _all, mv(-1) ;
rename bhid hid ; rename bpno pno ; sort hid pno ; save temp1, replace ;
use bhpsb1 ; keep pid hid pno doim doiy ; sort hid pno ; merge hid pno using temp1 ;
tab _merge ; keep if _merge==3 ; drop _merge ;
replace doiy=doiy-1900 if doiy>1000 & doiy~=. ; gen idate=(doiy*12)+doim ;
save blifemst, replace ;

/* reading clifejob */
use chid cpno cljseq cljsemp cljbgm cljbgy cljlfty cljlftm using
m:\bhps11\data\clifejob ;
quietly mvdecode _all, mv(-9) ; quietly mvdecode _all, mv(-8) ;
quietly mvdecode _all, mv(-7) ; quietly mvdecode _all, mv(-6) ;
quietly mvdecode _all, mv(-5) ; quietly mvdecode _all, mv(-4) ;
quietly mvdecode _all, mv(-3) ; quietly mvdecode _all, mv(-2) ;
quietly mvdecode _all, mv(-1) ;
rename chid hid ; rename cpno pno ; sort hid pno ; save temp2, replace ;
use bhpsc1 ; keep pid hid pno ; sort hid pno ; merge hid pno using temp2 ;
tab _merge ; keep if _merge==3 ; drop _merge ; save clifejob, replace ;

/* reading bindresp for date first left full-time education */
use pid bledendm bledendy using m:\bhps11\data\bindresp ;
quietly mvdecode _all, mv(-9) ; quietly mvdecode _all, mv(-8) ;
quietly mvdecode _all, mv(-7) ; quietly mvdecode _all, mv(-6) ;
quietly mvdecode _all, mv(-5) ; quietly mvdecode _all, mv(-4) ;
quietly mvdecode _all, mv(-3) ; quietly mvdecode _all, mv(-2) ;
quietly mvdecode _all, mv(-1) ;
sort pid ; save educend, replace ;

end ;


************************* ACorrect ******************************** ;
/* Corrections to Wave A */

capture program drop ACorrect ;
program define ACorrect ;

use bhpsa1 ;
replace jbstat=99 if jbstat==8 ; replace jbstat=8 if jbstat==7 ;
replace jbstat=7 if jbstat==6 ; replace jbstat=6 if jbstat==5 ;
replace jbstat=5 if jbstat==99 ;
gen doiy=91 ;
gen nemst=. ;
save bhpsa1, replace ; drop _all ;

use bhpsa2 ;
replace jhstat=99 if jhstat==8 ; replace jhstat=8 if jhstat==7 ;
replace jhstat=7 if jhstat==6 ; replace jhstat=6 if jhstat==5 ;
replace jhstat=5 if jhstat==99 ;
save bhpsa2, replace ; drop _all ;

end ;
```

```
************************* WSpells ********************************* ;
/* Create spells for wave data */

capture program drop WSpells ;
program define WSpells ;

/* append indresp */
local let1 "a" ; local let2 "b" ; local let3 "c" ; local let4 "d" ;
local let5 "e" ; local let6 "f" ; local let7 "g" ; local let8 "h" ;
local let9 "i" ; local let10 "j" ; local let11 "k" ;

local i=1 ;
use bhpsa1 ; save bhps1, replace ; clear ;
/* loop around waves b to k */
while `i'<=10 { ;
       local i=`i'+1 ;
       use bhps1 ;
       append using bhps`let`i''1 ;
       save bhps1, replace ;
       drop _all ;
       } ;

use bhps1 ;
/* dates */
replace doiy=doiy4 if doiy4~=. ; drop doiy4 ;
replace jbbgy=jbbgy4 if jbbgy4~=. ; drop jbbgy4 ;
replace jsbgy=jsbgy4 if jsbgy4~=. ; drop jsbgy4 ;
replace cjsbgy=cjsbgy4 if cjsbgy4~=. ; drop cjsbgy4 ;
replace doiy=doiy-1900 if doiy>1000 & doiy~=. ;
replace cjsbgy=cjsbgy-1900 if cjsbgy>1000 & cjsbgy~=. ;
replace jsbgy=jsbgy-1900 if jsbgy>1000 & jsbgy~=. ;
replace jbbgy=jbbgy-1900 if jbbgy>1000 & jbbgy~=. ;
replace doim=. if doim>12 ;
gen idate=(doiy*12)+doim ;
gen cjsflag=cjsbgm-12 if cjsbgm>12 ;
replace cjsflag=5 if cjsbgm==. ;
replace cjsbgm=1 if cjsbgm==13 ; replace cjsbgm=4 if cjsbgm==14 ;
replace cjsbgm=7 if cjsbgm==15 ; replace cjsbgm=10 if cjsbgm==16 ;
replace cjsbgm=7 if cjsbgm==. & cjsbgy<90 ;
gen cjsdate=(cjsbgy*12)+cjsbgm ;
gen jsflag=jsbgm-12 if jsbgm>12 ;
replace jsflag=5 if jsbgm==. ;
replace jsbgm=1 if jsbgm==13 ; replace jsbgm=4 if jsbgm==14 ;
replace jsbgm=7 if jsbgm==15 ; replace jsbgm=10 if jsbgm==16 ;
replace jsbgm=7 if jsbgm==. & jsbgy<90 ;
gen jsdate=(jsbgy*12)+jsbgm ;
gen jbflag=jbbgm-12 if jbbgm>12 ;
replace jbflag=5 if jbbgm==. ;
replace jbbgm=1 if jbbgm==13 ; replace jbbgm=4 if jbbgm==14 ;
replace jbbgm=7 if jbbgm==15 ; replace jbbgm=10 if jbbgm==16 ;
replace jbbgm=7 if jbbgm==. & jbbgy<90 ;
gen jbdate=(jbbgy*12)+jbbgm ;

/* status */
gen work=0 if (jbhas==2 & (jboff==2 | jboff==3)) ;
replace work=1 if jbsemp==2 ;
replace work=2 if jbsemp==1 ;
replace work=3 if (jbhas==1 | jboff==1) & jbsemp==. ;
replace work=99 if work==. ;
gen cstat=jbstat ;
replace cstat=(jbstat*100)+nemst if jbstat~=nemst & nemst~=. ;
replace cstat=nemst if jbstat==. ;
replace cstat=99 if cstat==. ;

/* Note: error in codebook: starting date for work appears to have been
   copied over to starting day for current spell where missing */
gen wdate=jsdate if work==1 ;
gen wflag=jsflag if work==1 ;
replace wdate=jbdate if work==2 ;
replace wflag=jbflag if work==2 ;
replace cjsdate=. if (work==1 | work==2) & cjsdate==wdate & cstat~=7 ;

/* Selecting the spell corresponding to the jobhist spells */
gen stat=work if work==1 | work==2 ;
```

```
replace stat=cstat if work==0 | work==99 | jbstat==7 ;
replace stat=cstat if (cstat==1 | cstat==2) & work==3 ;
replace stat=99 if stat==. ;
gen sdate=wdate if (work==1 | work==2) ;
replace sdate=cjsdate if work==0 | work==99 | jbstat==7 ;
replace sdate=cjsdate if (cstat==1 | cstat==2) & work==3 ;
gen sflag=wflag if work==1 | work==2 ;
replace sflag=cjsflag if work==0 | work==99 | jbstat==7 ;
replace sflag=cjsflag if (cstat==1 | cstat==2) & work==3 ;
save temp3, replace ;

/* spells1 */
use temp3 ; gen order=10 ; gen ref=wave ; rename idate edate ;
drop if sdate>edate & sdate~=. ;
keep pid wave ref order stat sdate sflag edate ;
sort pid wave ; save spells1, replace ;

/* additional information */
use temp3 ;
gen xflag=1 if work==1 & jsdate~=cjsdate & cjsdate~=. ;
replace xflag=1 if work==2 & jbdate~=cjsdate & cjsdate~=. ;
replace xflag=1 if work==3 & (cstat==1 | cstat==2 | cjsdate~=.) ;
replace xflag=2 if (work==1 | work==2) & (cstat~=1 & cstat~=2 & cstat~=99) ;
replace xflag=3 if jbstat==7 & (work==1 | work==2 | work==3) ;
keep pid wave idate cstat work cjsdate cjsflag wdate wflag jboff jboffy jbhas xflag ;
sort pid wave ; save tempadd, replace ;

/* append jobhist */
local let1 "a" ; local let2 "b" ; local let3 "c" ; local let4 "d" ;
local let5 "e" ; local let6 "f" ; local let7 "g" ; local let8 "h" ;
local let9 "i" ; local let10 "j" ; local let11 "k" ;

local i=1 ;
use bhpsa2 ; save bhps2, replace ; clear ;
/* loop around waves b to k */
while `i'<=10 { ;
       local i=`i'+1 ;
       use bhps2 ;
       append using bhps`let`i''2 ;
       save bhps2, replace ;
       drop _all ;
       } ;

use bhps2 ;
drop if jspno==. ;
drop _merge ;

/* status */
gen stat=jhstat ;
replace stat=1.5 if jhstat==1 ;
replace stat=1 if jhstat==2 & jhsemp==2 & wave==1 ;
replace stat=1 if jhstat==2 & jhsemp==3 & wave>1 ;
replace stat=99 if stat==. ;
replace jhbgy=jhbgy-1900 if jhbgy>1000 & jhbgy~=. ;

/* dates */
gen sflag=jhbgm-12 if jhbgm>12 ;
replace sflag=5 if jhbgm==. ;
replace jhbgm=1 if jhbgm==13 ; replace jhbgm=4 if jhbgm==14 ;
replace jhbgm=7 if jhbgm==15 ; replace jhbgm=10 if jhbgm==16 ;
replace jhbgm=7 if jhbgm==. & jhbgy<90 ;
replace jhbgy=jhbgy4 if jhbgy4~=. ; drop jhbgy4 ;
replace jhbgy=jhbgy-1900 if jhbgy>1000 & jhbgy~=. ;
gen sdate=(jhbgy*12)+jhbgm ;

/* spells2 */
gen order=10-jspno ; gen ref=wave+(jspno/10) ; sort pid wave ; save temp1, replace ;
use tempadd ; keep pid wave idate ; sort pid wave ; merge pid wave using temp1 ;
tab _merge ; drop if _merge==1 ; drop _merge ; drop if sdate>idate & sdate~=. ;
keep pid wave order stat sdate sflag ref ; save spells2, replace ;

/* append indresp and jobhist */
use spells1 ; append using spells2 ; gen count=1 ;
sort pid wave order ; quietly by pid wave: gen spell=sum(count) ;
sort pid wave ; quietly egen total=sum(count), by(pid wave) ; drop count order ;
reshape wide stat sdate sflag edate ref, i(pid wave total) j(spell) ;
```

```
sort pid wave ; save spells, replace ; drop _all ;

end ;


************************* EmpMerge ******************************** ;
/* Merges different job spells with same employer within wave together */
/* Note: cannot check whether there are any changes in job characteristics,
   including between full and part time because routed out of job questions
   if jhstat=1 */

capture program drop EmpMerge ;
program define EmpMerge ;

use spells ;

egen size=max(total) ;
gen promote=0 ;
local i=0 ;
/* loop around spells 1 to size-1 */
while `i'<=(size-1) { ;
       local i = `i'+1 ;
       quietly replace promote=promote+1 if stat`i'==1.5 ;
       } ;
save temp4, replace ;

keep if promote>0 ;
local i=0 ; local j=0 ;
/* loop around spells 1 to size-2 */
while `i'<=(size-2) { ;
       local i = `i'+1 ;
       local j = `i'+1 ;
       quietly gen omit`i'=1 if stat`i'==1.5 & (stat`j'==2 | stat`j'==1.5) ;
       quietly replace sdate`j'=sdate`i' if omit`i'==1 & sdate`j'>sdate`i' ;
       } ;
drop size ;
reshape long stat sdate sflag edate ref omit, i(pid wave total) j(spell) ;
drop if stat==. ; drop if omit==1 ;
rename spell order ; drop total ; gen count=1 ;
sort pid wave order ; quietly by pid wave: gen spell=sum(count) ;
sort pid wave ; quietly egen total=sum(count), by(pid wave) ;
drop count order omit ;
reshape wide stat sdate sflag edate ref, i(pid wave total) j(spell) ;
save temp5, replace ;

use temp4 ; keep if promote==0 ; append using temp5 ; sort pid wave ;
keep pid wave total stat* sdate* sflag* edate* ref* ; save wspells, replace ;

end ;


************************* BLife ********************************** ;

capture program drop BLife ;
program define BLife ;

use blifemst ;
gen ref=bleshno/100 ;
gen count=1 ; sort pid ; quietly egen total=sum(count), by(pid) ;
/* adjusting error in bleshno - ref do not run consecutively */
sort pid ref ; quietly by pid: gen spell=sum(count) ;

/* status */
gen stat=bleshst ; replace stat=stat-1 if stat>=3 & stat<=10 ;
replace stat=10 if bleshst==12 ; replace stat=99 if stat==. ;

/* dates */
gen sflag=bleshsm-12 if bleshsm>12 ;
replace sflag=5 if bleshsm==. ;
replace bleshsm=1 if bleshsm==13 ; replace bleshsm=4 if bleshsm==14 ;
replace bleshsm=7 if bleshsm==15 ; replace bleshsm=10 if bleshsm==16 ;
replace bleshsm=7 if bleshsm==. & bleshsy<90 ;
gen sdate=(bleshsy*12)+bleshsm ;
replace sdate=. if ref>0.01 ; replace sflag=. if ref>0.01 ;
gen eflag=bleshem-12 if bleshem>12 ;
replace eflag=5 if bleshem==. ;
```

```
replace bleshem=1 if bleshem==13 ; replace bleshem=4 if bleshem==14 ;
replace bleshem=7 if bleshem==15 ; replace bleshem=10 if bleshem==16 ;
replace bleshem=7 if bleshem==. ;
gen edate=(bleshey*12)+bleshem ;
replace edate=idate if bleshne==1 & edate==. ;

keep pid total stat sdate sflag edate eflag ref spell ;
reshape wide stat sdate sflag edate eflag ref, i (pid total) j(spell) ;
sort pid ; save blife, replace ;

end ;


************************* CLife ************************************ ;

capture program drop CLife ;
program define CLife ;

use clifejob ;
generate ref=cljseq/100 ;
gen count=1 ; sort pid ; quietly egen total=sum(count), by(pid) ;
/* allowing for possible errors in cljseq */
sort pid ref ; quietly by pid: gen spell=sum(count) ;

/* status */
gen stat=cljsemp ; replace stat=2 if cljsemp==2 | cljsemp==3 ; replace stat=99 if
stat==. ;

/* dates */
gen sflag=cljbgm-12 if cljbgm>12 ;
replace sflag=5 if cljbgm==. ;
replace cljbgm=1 if cljbgm==13 ; replace cljbgm=4 if cljbgm==14 ;
replace cljbgm=7 if cljbgm==15 ; replace cljbgm=10 if cljbgm==16 ;
replace cljbgm=7 if cljbgm==. & cljbgy<90 ;
gen sdate=(cljbgy*12)+cljbgm ;
gen eflag=cljlftm-12 if cljlftm>12 ;
replace eflag=5 if cljlftm==. ;
replace cljlftm=1 if cljlftm==13 ; replace cljlftm=4 if cljlftm==14 ;
replace cljlftm=7 if cljlftm==15 ; replace cljlftm=10 if cljlftm==16 ;
replace cljlftm=7 if cljlftm==. & cljlfty<90 ;
gen edate=(cljlfty*12)+cljlftm ;
replace edate=1088 if edate==. & sdate<1088 & spell==total ;

keep pid total stat sdate sflag edate eflag ref spell ;
reshape wide stat sdate sflag edate eflag ref, i (pid total) j(spell);
sort pid ; save clife, replace ;

end ;


************************* Demogs *********************************** ;

capture program drop Demogs ;
program define Demogs ;

use bhps1 ; keep pid wave sex age dobm doby scend feend qfedhi ; sort pid ;
save temp1, replace ;
merge pid using educend ; tab _merge ; drop _merge ; sort pid wave ;
save temp2, replace ;
merge pid wave using tempadd ; tab _merge ; drop _merge ; save temp3, replace ;

replace doby=doby-1900 ; gen temp1=max(scend,feend) ;
gen left1=((doby+temp1)*12) + dobm ;
gen left2=idate-((age-15)*12) if age>33 & (qfedhi==5 | qfedhi==8 | qfedhi>=10) ;
replace left2=idate-((age-16)*12) if age<34 & age>15
       & (qfedhi==5 | qfedhi==8 | qfedhi>=10) ;
replace left2=idate-((age-16)*12) if age>15 & (qfedhi==7 | qfedhi==9) ;
replace left2=idate-((age-17)*12) if age>16 & (qfedhi==3 | qfedhi==4) ;
replace left2=idate-((age-18)*12) if age>17 & (qfedhi==6) ;
replace left2=idate-((age-21)*12) if age>20 & (qfedhi==2) ;
replace left2=idate-((age-22)*12) if age>21 & (qfedhi==1) ;
replace bledendm=1 if bledendm==13 ; replace bledendm=4 if bledendm==14 ;
replace bledendm=7 if bledendm==15 ; replace bledendm=10 if bledendm==16 ;
replace bledendm=7 if bledendm==. & bledendy<90 ; gen left3=(bledendy*12)+bledendm ;
gen left=max(left1,left2,left3) ;
drop temp* dobm doby scend feend qfedhi bledendm bledendy left1 left2 left3 ;
```

```
drop if wave==. ; sort pid wave ; save addinfo, replace ;

end ;


/*******************************************************************************
       Running the Show
*******************************************************************************/

ReadData ;
ACorrect ;
WSpells ;
EmpMerge ;
BLife ;
CLife ;
Demogs ;


log close ;
set more 0 ;
#delimit cr



*******************************************************************************************
*******************************************************************************************
*******************************************************************************************
/* bhpshist2: second step in creating the basic data */

#delimit ;
set more 1 ;
cd c:\gillian\disk10ext ;
log using bhpshist2.log, replace ;


************************** BAmend ********************************** ;
/* Amends Blife: adds start dates, converts stat to 5 categories and
   merges consecutive spells of the same state */

capture program drop BAmend ;
program define BAmend ;

use blife ;
/* loop around spells 1 to 38 */
gen sdate39=. ; gen stat39=. ; gen sflag39=. ;
local i=0 ;
local j=0 ;
while `i' <= 37 { ;
       local i = `i'+1 ;
       local j = `i'+1 ;
       quietly replace sdate`j'=edate`i' if stat`j'~=. ;
       quietly replace sflag`j'=eflag`i' if stat`j'~=. ;
       } ;
drop sdate39 stat39 sflag39 ;
save temp1, replace ;

use temp1 ; keep if total<8 ;
drop total stat8-stat38 sdate8-sdate38 sflag8-sflag38 edate8-edate38
     eflag8-eflag38 ref8-ref38 ;
/* loop around spells 1 to 8 */
local i=0 ;
local j=0 ;
while `i' <= 6 { ;
       local i = `i'+1 ;
       local j = `i'+1 ;
       quietly replace stat`i'=4 if stat`i'>3 & stat`i'<12 & stat`i'~=7
                      & stat`i'~=5 & stat`i'~=9 ;
       quietly replace stat`i'=2 if stat`i'==5 ;
       quietly replace stat`i'=2 if stat`i'==9 ;
       } ;
/* loop around spells 1 to 8 */
gen stat8=. ; gen sdate8=. ; gen sflag8=. ; gen ref8=. ;
local i=0 ;
local j=0 ;
while `i' <= 6 { ;
       local i = `i'+1 ;
```

```
        local j = `i'+1 ;
        quietly gen omit`i'=1 if stat`i'==stat`j' & sdate`i'<=sdate`j' ;
        quietly replace sdate`j'=sdate`i' if omit`i'==1 ;
        quietly replace sflag`j'=sflag`i' if omit`i'==1 ;
        quietly replace ref`j'=(ref`j'/100)+ref`i' if omit`i'==1 ;
        } ;
drop stat8 sdate8 sflag8 ref8 ;
reshape long stat sdate sflag edate eflag ref omit, i(pid) j(spell) ;
drop if stat==. ; drop if omit==1 ; drop spell ;
gen count=1 ; sort pid ; quietly by pid: gen spell=sum(count) ;
sort pid ; egen total=sum(count), by(pid) ; drop omit count ;
save temp2, replace ;

use temp1 ; keep if total>7 ;
/* loop around spells 1 to 38 */
local i=0 ;
local j=0 ;
while `i' <= 37 { ;
        local i = `i'+1 ;
        local j = `i'+1 ;
        quietly replace stat`i'=4 if stat`i'>3 & stat`i'<12 & stat`i'~=7
                        & stat`i'~=5 & stat`i'~=9 ;
        quietly replace stat`i'=2 if stat`i'==5 ;
        quietly replace stat`i'=2 if stat`i'==9 ;
        } ;
/* loop around spells 1 to 38 */
gen stat39=. ; gen sdate39=. ; gen sflag39=. ; gen ref39=. ;
local i=0 ;
local j=0 ;
while `i' <= 37 { ;
        local i = `i'+1 ;
        local j = `i'+1 ;
        quietly gen omit`i'=1 if stat`i'==stat`j' & sdate`i'<=sdate`j' ;
        quietly replace sdate`j'=sdate`i' if omit`i'==1 ;
        quietly replace sflag`j'=sflag`i' if omit`i'==1 ;
        quietly replace ref`j'=(ref`j'/100)+ref`i' if omit`i'==1 ;
        } ;
drop stat39 sdate39 sflag39 ref39 ;
reshape long stat sdate sflag edate eflag ref omit, i(pid total) j(spell) ;
drop if stat==. ; drop if omit==1 ; drop spell ;
gen count=1 ; sort pid ; quietly by pid: gen spell=sum(count) ;
drop total ; egen total=sum(count), by(pid) ; drop omit count ;
save temp3, replace ;

use temp2 ; append using temp3 ; sort pid ; save tempb, replace ;

end ;


************************** CAmend ********************************* ;
/* Amends Clife: reshapes long */

capture program drop CAmend ;
program define CAmend ;

use clife ;
reshape long stat sdate sflag edate eflag ref, i(pid total) j(spell) ;
drop if stat==. ; sort pid ; save tempc, replace ;

end ;


************************** BCAmend ******************************** ;
/* Creates BCtemp by imposing employer-to-employer moves
   from clifejob on blifemst */

capture program drop BCAmend ;
program define BCAmend ;

use clife ;
/* loop around spells 1 to 21 */
local i=0 ; local j=0 ;
while `i' <= 19 { ;
        local i = `i'+1 ;
        local j = `i'+1 ;
        quietly gen jsdate`j'=sdate`j' if ((sdate`j'-edate`i')<2) & stat`i'==2
```

```
                                        & stat`j'==2 ;
        quietly gen jsflag`j'=sflag`j' if ((sdate`j'-edate`i')<2) & stat`i'==2
                                        & stat`j'==2 ;
        } ;
keep pid jsdate* jsflag* ; reshape long jsdate jsflag, i(pid) j(spell) ;
keep if jsdate~=. ;
gen count=1 ; sort pid spell ; quietly by pid: gen move=sum(count) ;
drop spell count ;
reshape wide jsdate jsflag, i(pid) j(move) ; sort pid ; save temp1, replace ;

use clife ; keep pid ; sort pid ; merge pid using tempb ; tab _merge ;
gen count=1 if stat==2 ; replace count=. if sdate>1088 & sdate~=. ; sort pid ;
egen totemp=sum(count), by(pid) ; keep if (_merge==3 | totemp==0) & stat~=. ;
drop _merge totemp count ; sort pid ; merge pid using temp1 ; tab _merge ;
drop if _merge==2 ; drop _merge ;
/* loop around moves 1 to 17 */
local i=0 ;
while `i' <= 16 { ;
        local i = `i'+1 ;
        quietly gen match`i'=1 if sdate<jsdate`i' & edate>jsdate`i' & jsdate`i'~=.
                                & edate~=. & stat==2 ;
        sort pid ; egen check`i'=sum(match`i'), by(pid) ;
        } ;
egen stot=rsum(match*) ; tab stot ; sort pid ; egen ptot=sum(stot), by(pid) ;
tab ptot ; save checkbc, replace ;
use checkbc ; keep if ptot==0 ; drop match* jsdate* jsflag* check* stot ptot ;
save temp3, replace ;
use checkbc ; keep if ptot>0 ; drop check* stot ptot ; save temp4, replace ;

use temp4 ;
gen match0=1 ; gen jsdate0=. ; gen jsflag0=. ;
reshape long jsdate jsflag match, i(pid total stat sdate sflag edate eflag ref spell)
                                j(move) ;
keep if match==1 ; drop match ; rename move temp ; gen count=1 ; sort pid spell temp ;
quietly by pid spell: gen move=sum(count) ; drop temp count ;
reshape wide jsdate jsflag, i(pid total stat sdate sflag edate eflag ref spell)
j(move) ;
/* loop around moves 1 to 18 */
local i=0 ; local j=0 ;
while `i' <= 16 { ;
        local i = `i'+1 ; local j = `i'+1 ;
        quietly gen jedate`i'=jsdate`j' ;
        quietly gen jeflag`i'=jsflag`j' ;
        } ;
reshape long jsdate jsflag jedate jeflag,
            i(pid total stat sdate sflag edate eflag ref spell) j(move) ;
drop if jsdate==. & jedate==. & move>1 ;
quietly replace sdate=jsdate if jsdate~=. ;
quietly replace sflag=jsflag if jsdate~=. ;
quietly replace edate=jedate if jedate~=. ;
quietly replace eflag=jeflag if jedate~=. ;
drop jsdate jsflag jedate jeflag total ;
rename spell temp ; gen count=1 ; sort pid temp move ;
quietly by pid: gen spell=sum(count) ; egen total=sum(count), by(pid) ;
drop move count temp ; save temp5, replace ;

use temp5 ; append using temp3 ; save tempbc, replace ;

end ;


************************* WMain ******************************** ;
/* WMain: corrects for main activity */

capture program drop WMain ;
program define WMain ;

use addinfo ; keep if xflag==2 ; keep pid wave cstat cjsdate cjsflag xflag ;
rename wave ref ; sort pid ref ; save temp1, replace ;
use wspells ; keep pid wave stat* sdate* sflag* ref* ;
reshape long stat sdate sflag ref, i(pid wave) j(spell) ; drop if stat==. ;
sort pid ref ; merge pid ref using temp1 ; tab _merge ; drop _merge ;
replace stat=cstat if xflag==2 ; replace sdate=. if xflag==2 ;
replace sflag=. if xflag==2 ; drop cstat cjsdate cjsflag xflag ;
reshape wide stat sdate sflag ref, i(pid wave) j(spell) ;
sort pid wave ; save temp2, replace ;
```

```
use wspells ; drop stat* sdate* sflag* ref* ; sort pid wave ;
merge pid wave using temp2 ; tab _merge ; drop _merge ;
save tempmain, replace ;

end ;


************************* WWork ******************************** ;
/* WWork: corrects for all work */

capture program drop WWork ;
program define WWork;

use addinfo ; keep if xflag==3 ; keep pid wave work wdate wflag xflag ;
rename wave ref ; sort pid ref ; save temp1, replace ;
use wspells ; keep pid wave stat* sdate* sflag* ref* ;
reshape long stat sdate sflag ref, i(pid wave) j(spell) ; drop if stat==. ;
sort pid ref ; merge pid ref using temp1 ; tab _merge ; drop _merge ;
replace stat=work if xflag==3 ; replace sdate=. if xflag==3 ;
replace sflag=. if xflag==3 ; drop work wdate wflag ;
rename spell temp ; gen count=1 ; sort pid wave temp ;
quietly by pid wave: gen spell=sum(count) ; drop temp count ;
reshape wide stat sdate sflag ref xflag, i(pid wave) j(spell) ;
sort pid wave ; save temp2, replace ;
use wspells ; drop stat* sdate* sflag* ref* ; sort pid wave ;
merge pid wave using temp2 ; tab _merge ; drop _merge ;
save tempwork, replace ;

end ;


************************* WAmend ******************************** ;
/* Amends Wspells: adds end dates, converts stat to 5 categories,
   merges consecutive spells of the same state other than employment */

capture program drop WAmend ;
program define WAmend ;

egen size=max(total) ; local s=(size+1) ;
/* loop around spells 1 to size-1 */
gen sdate`s'=. ; gen sflag`s'=. ;
local i=0 ;
local j=0 ;
while `i' <= (size-1) { ;
       local i = `i'+1 ;
       local j = `i'+1 ;
       quietly gen eflag`i'=sflag`j' if edate`i'==. & stat`i'~=. ;
       quietly replace edate`i'=sdate`j' if edate`i'==. & stat`i'~=. ;
       } ;
drop sdate`s' sflag`s' ;
/* loop around spells 1 to (size-1) */
local i=0 ;
local j=0 ;
while `i' <= (size-1) { ;
       local i = `i'+1 ;
       local j = `i'+1 ;
       quietly replace stat`i'=2 if stat`i'==1.5 | stat`i'==2 ;
       quietly replace stat`i'=2 if stat`i'==5 | stat`i'==105 | stat`i'==205
              | stat`i'==305 | stat`i'==405 | stat`i'==605 | stat`i'==705
              | stat`i'==805 | stat`i'==905 | stat`i'==1005 ;
       quietly replace stat`i'=2 if stat`i'==9 | stat`i'==109 | stat`i'==209
              | stat`i'==309 | stat`i'==409 | stat`i'==509 | stat`i'==609
              | stat`i'==709 | stat`i'==809 | stat`i'==1009 ;
       quietly replace stat`i'=3 if stat`i'==3 | stat`i'==103 | stat`i'==203
              | stat`i'==403 | stat`i'==503 | stat`i'==603 | stat`i'==703
              | stat`i'==803 | stat`i'==903 | stat`i'==1003 ;
       quietly replace stat`i'=7 if stat`i'==7 | stat`i'==107 | stat`i'==207
              | stat`i'==307 | stat`i'==407 | stat`i'==507 | stat`i'==607
              | stat`i'==807 | stat`i'==907 | stat`i'==1007 ;
       quietly replace stat`i'=4 if stat`i'>3 & stat`i'~=7 & stat`i'~=99
              & stat`i'~=. ;
       } ;
/* loop around spells 1 to size-1 */
gen sdate`s'=. ; gen stat`s'=. ; gen edate`s'=. ; gen ref`s'=. ;
local i=0 ;
local j=0 ;
```

```
while `i' <= (size-1) { ;
       local i = `i'+1 ;
       local j = `i'+1 ;
       quietly gen omit`i'=1 if stat`i'==stat`j' & stat`i'~=2 & stat`i'~=99
                              & sdate`i'<=sdate`j' ;
       quietly replace sdate`j'=sdate`i' if omit`i'==1 ;
       quietly replace ref`i'=(ref`i'-wave)/10 if omit`i'==1 ;
       quietly replace ref`j'=ref`j'+ref`i' if omit`i'==1 ;
       } ;
drop sdate`s' stat`s' edate`s' ref`s' size ;
reshape long stat sdate sflag edate eflag ref omit, i(pid wave total) j(spell) ;
drop if stat==. ; drop if omit==1 ; drop spell ;
gen count=1 ; sort pid wave ; quietly by pid wave: gen spell=sum(count) ;
drop total ; egen total=sum(count), by(pid wave) ; drop omit count ;

end ;


********************************** BPanel ******************************** ;
/* Creates BPanel */

capture program drop BPanel ;
program define BPanel ;

use tempb ; gen wave=0 ; append using wavemain ; sort pid ;
save temp1, replace ;
use addinfo ; keep pid wave idate ; reshape wide idate, i(pid) j(wave) ;
sort pid ; save tempdate, replace ; merge pid using temp1 ; tab _merge ;
drop _merge ;
sort pid ; egen wavemin=min(wave), by(pid) ; gen idate0=1088 if wavemin==0 ;
drop wavemin ;
drop if wave==0 & sdate>idate0 & sdate~=. & idate1~=. ;
drop if wave==0 & sdate>1100 & sdate~=. ;
replace edate=idate0 if wave==0 & edate>idate0 & edate~=. & idate1~=. ;
replace edate=1100 if wave==0 & edate>1100 & edate~=. & idate1==. ;
rename spell temp ; drop total ; gen count=1 ; sort pid wave temp ;
quietly by pid: gen spell=sum(count) ; sort pid ;
egen total=sum(count), by(pid) ; drop count temp ; save blpanel, replace ;
keep pid total spell wave stat sdate sflag edate eflag ;
reshape wide wave stat sdate sflag edate eflag, i(pid total) j(spell) ;
sort pid ; save temp2, replace ;
use blpanel ; keep if spell==1 ; keep pid idate* ; sort pid ;
merge pid using temp2 ; tab _merge ; drop _merge ; sort pid ;
save bpanel, replace ;

end ;


********************************** CPanel ******************************** ;
/* Creates CPanel */

capture program drop CPanel ;
program define CPanel ;

use tempc ; gen wave=0 ; append using wavework ; sort pid ;
save temp1, replace ;
use addinfo ; keep pid wave idate ; reshape wide idate, i(pid) j(wave) ;
sort pid ; save tempdate, replace ; merge pid using temp1 ; tab _merge ;
drop _merge ; sort pid ; egen wavemin=min(wave), by(pid) ;
gen idate0=1088 if wavemin==0 ; drop wavemin ;
drop if wave==0 & sdate>idate0 & sdate~=. & idate1~=. ;
drop if wave==0 & sdate>1100 & sdate~=. ;
replace edate=idate0 if wave==0 & edate>idate0 & edate~=. & idate1~=. ;
replace edate=1100 if wave==0 & edate>1100 & edate~=. & idate1==. ;
rename spell temp ; drop total ; gen count=1 ; sort pid wave temp ;
quietly by pid: gen spell=sum(count) ; sort pid ;
egen total=sum(count), by(pid) ; drop count temp ; save clpanel, replace ;
keep pid total spell wave stat sdate sflag edate eflag ;
reshape wide wave stat sdate sflag edate eflag, i(pid total) j(spell) ;
sort pid ; save temp2, replace ;
use clpanel ; keep if spell==1 ; keep pid idate* ; sort pid ;
merge pid using temp2 ; tab _merge ; drop _merge ; sort pid ;
save cpanel, replace ;

end ;
```

```
********************************** BCPanel ******************************** ;
/* Creates BCPanel */

capture program drop BCPanel ;
program define BCPanel ;

use tempbc ; gen wave=0 ; append using wavemain ; sort pid ;
save temp1, replace ;
use addinfo ; keep pid wave idate ; reshape wide idate, i(pid) j(wave) ;
sort pid ; save tempdate, replace ; merge pid using temp1 ; tab _merge ;
drop _merge ; sort pid ; egen wavemin=min(wave), by(pid) ;
gen idate0=1088 if wavemin==0 ; drop wavemin ;
drop if wave==0 & sdate>idate0 & sdate~=. & idate1~=. ;
drop if wave==0 & sdate>1100 & sdate~=. ;
replace edate=idate0 if wave==0 & edate>idate0 & edate~=. & idate1~=. ;
replace edate=1100 if wave==0 & edate>1100 & edate~=. & idate1==. ;
rename spell temp ; drop total ; gen count=1 ; sort pid wave temp ;
quietly by pid: gen spell=sum(count) ; sort pid ;
egen total=sum(count), by(pid) ; drop count temp ; save bclpanel, replace ;
keep pid total spell wave stat sdate sflag edate eflag ;
reshape wide wave stat sdate sflag edate eflag, i(pid total) j(spell) ;
sort pid ; save temp2, replace ;
use bclpanel ; keep if spell==1 ; keep pid idate* ; sort pid ;
merge pid using temp2 ; tab _merge ; drop _merge ; sort pid ;
save bcpanel, replace ;

end ;


/*****************************************************************************
                        RUNNING THE SHOW
*****************************************************************************/

BAmend ;
CAmend ;
BCAmend ;
WMain ;
WWork ;
use tempmain ; WAmend ; save wavemain, replace ;
use tempwork ; WAmend ; save wavework, replace ;
BPanel ;
CPanel ;
BCPanel ;


log close ;
set more 0 ;
#delimit cr


*******************************************************************************
*******************************************************************************
*******************************************************************************
/* bhpshist3: constructs the history data sets for the different methods */

#delimit ;
set more 1 ;
cd c:\gillian\disk10ext ;
log using bhpshist3.log, replace ;


************************** DateFixEM ****************************************;
/* DateFixEM: amends imputed dates to generate consistency for method C */

capture program drop DateFixEM ;
program define DateFixEM ;

/* correction for first spell in blife */
replace edate1=. if sdate1>edate1 & sdate1~=. & eflag1==5 & wave1==0 ;
replace sdate1=. if sdate1>edate1 & sdate1~=. & sflag1==5 & wave1==0 ;
replace edate1=edate1+1 if (sdate1-edate1==1) & (eflag1>=1 & eflag1<=4)
                         & wave1==0  ;
replace sdate1=sdate1-1 if (sdate1-edate1==1) & (sflag1>=1 & sflag1<=4)
                         & wave1==0 ;
replace edate1=edate1+11 if (sdate1-edate1>0 & sdate1-edate1<10)
```

```
                          & eflag1==1 & wave1==0 ;
replace sdate2=edate1 if wave1==0 & wave2==0 ;

egen size=max(total) ;
/* loop around spells 1 to size */
local i=0 ; local j=0 ;
while `i' <= size-2 { ;
       local i = `i'+1 ; local j = `i'+1 ;
       /* correction for blife spells */
       quietly replace edate`i'=. if edate`i'-edate`j'>0 & edate`j'~=.
                          & eflag`i'==5 & wave`j'==0 ;
       quietly replace edate`j'=. if edate`i'-edate`j'>0 & edate`i'~=.
                          & eflag`j'==5 & wave`j'==0 ;
       quietly replace edate`i'=edate`i'-1 if (edate`i'-edate`j'==1)
                          & (eflag`i'>=1 & eflag`i'<=4) & wave`j'==0 ;
       quietly replace edate`j'=edate`j'+1 if (edate`i'-edate`j'==1)
                          & (eflag`j'>=1 & eflag`j'<=4) & wave`j'==0 ;
       quietly replace edate`j'=edate`j'+11 if
                           (edate`i'-edate`j'>0 & edate`i'-edate`j'<10)
                           & eflag`j'==1 & wave`j'==0 ;
       quietly replace sdate`j'=edate`i' if wave`j'==0 ;
       /* correction for within wave spells */
       quietly gen do`i'=1 if wave`i'>0 & wave`i'~=. & wave`i'==wave`j' ;
       quietly replace sdate`i'=. if sdate`i'-sdate`j'>0 & sdate`j'~=.
                          & sflag`i'==5 & do`i'==1 ;
       quietly replace sdate`j'=. if sdate`i'-sdate`j'>0 & sdate`i'~=.
                          & sflag`j'==5 & do`i'==1 ;
       quietly replace sdate`i'=sdate`i'-1 if (sdate`i'-sdate`j'==1)
                          & (sflag`i'>=1 & sflag`i'<=4) & do`i'==1 ;
       quietly replace sdate`j'=sdate`j'+1 if (sdate`i'-sdate`j'==1)
                          & (sflag`j'>=1 & sflag`j'<=4) & do`i'==1 ;
       quietly replace sdate`j'=sdate`j'+11 if
                          (sdate`i'-sdate`j'>0 & sdate`i'-sdate`j'<10)
                          & sflag`j'==1 & do`i'==1 ;
       quietly replace edate`i'=sdate`j' if do`i'==1 ;
       } ;
drop sflag* eflag* do* size ;

end ;


************************** DateFixW ****************************************;
/* DateFixW: amends imputed dates to generate consistency for method C*/

capture program drop DateFixW ;
program define DateFixW ;

/* correction for first spell in clife */
replace edate1=. if sdate1>edate1 & sdate1~=. & eflag1==5 & wave1==0 ;
replace sdate1=. if sdate1>edate1 & sdate1~=. & sflag1==5 & wave1==0 ;
replace edate1=edate1+1 if (sdate1-edate1==1) & (eflag1>=1 & eflag1<=4)
                         & wave1==0  ;
replace sdate1=sdate1-1 if (sdate1-edate1==1) & (sflag1>=1 & sflag1<=4)
                         & wave1==0 ;
replace edate1=edate1+11 if (sdate1-edate1>0 & sdate1-edate1<10) & eflag1==1
                         & wave1==0 ;
replace sdate2=edate1 if wave1==0 & wave2==0 ;

egen size=max(total) ; local s = size+1 ;
/* loop around spells 1 to size */
local i=0 ; local j=0 ; local k=0 ;
gen wave`s'=. ; gen sdate`s'=. ; gen edate`s'=. ; gen sflag`s'=. ;
gen eflag`s'=. ;
while `i' <= size-2 { ;
       local i = `i'+1 ; local j = `i'+1 ; local k = `i'+2 ;
       /* correction for clife spells */
       quietly replace sdate`i'=. if sdate`i'>edate`i' & sflag`i'==5
                          & wave`j'==0 ;
       quietly replace sdate`j'=. if edate`i'>sdate`j' & edate`i'~=.
                          & sflag`j'==5 & wave`j'==0 ;
       quietly replace edate`i'=. if sdate`i'>edate`i' & sdate`i'~=.
                          & eflag`i'==5 & wave`j'==0 ;
       quietly replace edate`i'=. if edate`i'>sdate`j' & eflag`i'==5
                          & wave`j'==0 ;
       quietly replace sdate`i'=sdate`i'-1 if (sdate`i'-edate`i'==1)
                          & (sflag`i'<=1 & sflag`i'>=4) & wave`j'==0 ;
```

```
        quietly replace edate`i'=edate`i'+1 if (sdate`i'-edate`i'==1)
                              & (eflag`i'<=1 & eflag`i'>=4) & wave`j'==0 ;
        quietly replace edate`i'=edate`i'-1 if (edate`i'-sdate`j'==1)
                              & (eflag`i'<=1 & eflag`i'>=4) & wave`j'==0 ;
        quietly replace sdate`j'=sdate`j'+1 if (edate`i'-sdate`j'==1)
                              & (sflag`j'<=1 & sflag`j'>=4) & wave`j'==0 ;
        quietly replace edate`i'=edate`i'+11 if
               (sdate`i'-edate`i'>0 & sdate`i'-edate`i'<12 & sdate`j'-edate`i'>10)
               & eflag`i'==1 & wave`j'==0 ;
        quietly replace sdate`j'=sdate`j'+11 if
               (edate`i'-sdate`j'>0 & edate`i'-sdate`i'<12 & edate`j'-sdate`j'>10)
               & sflag`j'==1 & wave`j'==0 ;
        quietly gen pair`i'=1 if
               (sdate`i'-edate`i'>0 & sdate`i'-edate`i'<12) & eflag`i'==1 &
               (sdate`j'==edate`i' & edate`j'-sdate`j'>10) & sflag`j'==1
               & wave`j'==0 ;
        quietly replace edate`i'=edate`i'+11 if pair`i'==1 ;
        quietly replace sdate`j'=sdate`j'+11 if pair`i'==1 ;
        quietly drop pair`i' ;
        quietly gen pair`i'=1 if
               (edate`i'-sdate`j'>0 & edate`i'-sdate`j'<12) & sflag`j'==1 &
               (sdate`j'==edate`j' & sdate`k'-edate`j'>10) & eflag`j'==1
               & wave`j'==0 ;
        quietly replace sdate`j'=sdate`j'+11 if pair`i'==1 ;
        quietly replace edate`j'=edate`j'+11 if pair`i'==1 ;
        quietly drop pair`i' ;
        quietly gen triple`i'=1 if
               (edate`i'-sdate`j'>0 & edate`i'-sdate`j'<12) & sflag`j'==1 &
               (edate`j'==sdate`j') & eflag`j'==1 &
               (edate`j'==sdate`k' & edate`k'-sdate`k'>10) & sflag`k'==1
               & wave`k'==0 ;
        quietly replace sdate`j'=sdate`j'+11 if triple`i'==1 ;
        quietly replace edate`j'=edate`j'+11 if triple`i'==1 ;
        quietly replace sdate`k'=sdate`k'+11 if triple`i'==1 ;
        quietly drop triple`i' ;
        quietly gen triple`i'=1 if
               (sdate`i'-edate`i'>0 & sdate`i'-edate`i'<12) & eflag`i'==1 &
               (sdate`j'==edate`i') & sflag`j'==1 &
               (edate`j'==sdate`j' & sdate`k'-edate`j'>10) & eflag`j'==1
               & wave`k'==0 ;
        quietly replace edate`i'=edate`i'+11 if triple`i'==1 ;
        quietly replace sdate`j'=sdate`j'+11 if triple`i'==1 ;
        quietly replace edate`j'=edate`j'+11 if triple`i'==1 ;
        quietly drop triple`i' ;
        /* correction for within wave spells */
        quietly gen do`i'=1 if wave`i'>0 & wave`i'~=. & wave`i'==wave`j' ;
        quietly replace sdate`i'=. if sdate`i'-sdate`j'>0 & sdate`j'~=.
               & sflag`i'==5 & do`i'==1 ;
        quietly replace sdate`j'=. if sdate`i'-sdate`j'>0 & sdate`i'~=.
               & sflag`j'==5 & do`i'==1 ;
        quietly replace sdate`i'=sdate`i'-1 if (sdate`i'-sdate`j'==1)
               & (sflag`i'>=1 & sflag`i'<=4) & do`i'==1 ;
        quietly replace sdate`j'=sdate`j'+1 if (sdate`i'-sdate`j'==1)
               & (sflag`j'>=1 & sflag`j'<=4) & do`i'==1 ;
        quietly replace sdate`j'=sdate`j'+11 if
               (sdate`i'-sdate`j'>0 & sdate`i'-sdate`j'<10)
               & sflag`j'==1 & do`i'==1 ;
        quietly replace edate`i'=sdate`j' if do`i'==1 ;
        quietly drop do`i' ;
        } ;
drop wave`s' sdate`s' edate`s' sflag* eflag* size ;

end ;


************************** CrosWave ******************************************;
/* CrosWave: divides the data for SubCW */

capture program drop CrosWave ;
program define CrosWave ;

egen size=max(total) ; local s=size ;
save temp1, replace ;
use temp1 ; keep if total<10 ; gen subsize=9 ;
drop wave10-wave`s' stat10-stat`s' sdate10-sdate`s' edate10-edate`s' ;
SubCW ; save temp2, replace ;
```

```
use temp1 ; keep if total>=10 & total<16 ; gen subsize=15 ;
drop wave16-wave`s' stat16-stat`s' sdate16-sdate`s' edate16-edate`s' ;
SubCW ; save temp3, replace ;
use temp1 ; keep if total>=16 & total<25 ; gen subsize=24 ;
drop wave25-wave`s' stat25-stat`s' sdate25-sdate`s' edate25-edate`s' ;
SubCW ; save temp4, replace ;
use temp1 ; keep if total>=25 ; gen subsize=size ;
SubCW ; save temp5, replace ;
use temp2 ; append using temp3 ; append using temp4 ; append using temp5 ;
drop size ;

end ;


************************** SubCW ****************************************;
/* SubCW: main program for cross wave: drops spells which end before prior
          interview and merges matched spells across waves */

capture program drop SubCW ;
program define SubCW ;

save temp6, replace ; drop idate* ;
reshape long wave stat sdate edate, i(pid subsize) j(spell) ; drop if stat==. ;
sort pid ; save temp7, replace ;
use temp6 ; keep pid idate* ; sort pid ; merge pid using temp7 ;
tab _merge ; drop _merge ;
gen idate=idate0 if wave==0 ; replace idate=idate1 if wave==1 ;
replace idate=idate2 if wave==2 ; replace idate=idate3 if wave==3 ;
replace idate=idate4 if wave==4 ; replace idate=idate5 if wave==5 ;
replace idate=idate6 if wave==6 ; replace idate=idate7 if wave==7 ;
replace idate=idate8 if wave==8 ; replace idate=idate9 if wave==9 ;
gen pidate=idate0 if wave==1 ; replace pidate=idate1 if wave==2 ;
replace pidate=idate2 if wave==3 ; replace pidate=idate3 if wave==4 ;
replace pidate=idate4 if wave==5 ; replace pidate=idate5 if wave==6 ;
replace pidate=idate6 if wave==7 ; replace pidate=idate7 if wave==8 ;
replace pidate=idate8 if wave==9 ;
drop if edate<pidate & pidate~=. ; drop if sdate>idate & sdate~=. ;
drop total ; rename spell temp ; gen count=1 ; sort pid temp ;
quietly by pid: gen spell=sum(count) ; quietly egen total=sum(count), by(pid) ;
drop count temp pidate idate subsize ; quietly egen subsize=max(total) ;
save temp8, replace ;  drop idate* ;
reshape wide wave stat sdate edate, i(pid subsize) j(spell) ;
sort pid ; save temp9, replace ;
use temp8 ; keep pid spell idate* ; keep if spell==1 ; drop spell ; sort pid ;
merge pid using temp9 ; tab _merge ; drop _merge ;
/* loop around spells 1 to subsize */
local i=0 ; local j=0 ;
while `i' <= subsize-2 { ;
       local i = `i'+1 ; local j = `i'+1 ;
       quietly gen omit`i'=1 if wave`i'~=wave`j' & stat`i'==stat`j'
                     & stat`i'~=2 & sdate`j'<=edate`i' & edate`i'~=. ;
       quietly replace omit`i'=1 if wave`i'~=wave`j' & stat`i'==stat`j'
                     & stat`i'==2 & sdate`j'<edate`i' & edate`i'~=. ;
       quietly replace sdate`j'=sdate`i' if omit`i'==1 ;
       quietly replace wave`j'=wave`i' if omit`i'==1 ;
       } ;
drop subsize ; save temp10, replace ;
use temp10 ; keep pid idate* ; sort pid ; save temp11, replace ;
use temp10 ; drop idate* ;
reshape long wave stat sdate edate omit, i(pid) j(spell) ;
drop if stat==. ; drop if omit==1 ; drop total omit ; rename spell temp ;
gen count=1 ; sort pid temp ; quietly by pid: gen spell=sum(count) ;
quietly egen total=sum(count), by(pid) ;
drop temp count ; sort pid ; merge pid using temp11 ;
tab _merge ; drop _merge ;

end ;


************************** SeamsA ****************************************;
/* SeamsA: divides the data for SubSA */

capture program drop SeamsA ;
program define SeamsA ;

save temp1, replace ;
```

```
use temp1 ; keep if total<10 ; gen subsize=9 ; SubSA ;
save temp2, replace ;
use temp1 ; keep if total>=10 & total<16 ; gen subsize=15 ; SubSA ;
save temp3, replace ;
use temp1 ; keep if total>=16 & total<25 ; gen subsize=24 ; SubSA ;
save temp4, replace ;
use temp1 ; keep if total>=25 ; egen subsize=max(total) ; SubSA ;
save temp5, replace ;
use temp2 ; append using temp3 ; append using temp4 ; append using temp5 ;

end ;


************************** SubSA ****************************************;
/* SubSA: creates false spell breaks at seams for method A */

capture program drop SubSA ;
program define SubSA ;

save temp6, replace ; drop idate* ;
reshape wide wave stat sdate edate, i(pid subsize) j(spell) ;
sort pid ; save temp7, replace ;
use temp6 ; keep pid spell idate* ; keep if spell==1 ; drop spell ; sort pid ;
merge pid using temp7 ; tab _merge ; drop _merge ;
/* loop around spells 1 to subsize */
local i=0 ; local j=0 ;
while `i' <= subsize-2 { ;
       local i = `i'+1 ; local j = `i'+1 ;
       quietly gen seam`i'=1 if wave`i'~=wave`j'
                           & (stat`i'~=stat`j' | stat`i'==99 | stat`j'==99)
                           & sdate`j'<=edate`i' & edate`i'~=. ;
       quietly replace sdate`j'=edate`i' if seam`i'==1 & wave`i'~= 0 ;
       quietly replace sdate`j'=1088 if seam`i'==1 & wave`i'==0 & wave`j'==1
                           & sdate`j'<1088 ;
       quietly replace edate`i'=1088 if seam`i'==1 & wave`i'==0 & wave`j'==1
                           & sdate`j'<=1088 ;
       quietly replace edate`i'=sdate`j' if seam`i'==1 & wave`i'==0 & wave`j'==1
                           & sdate`j'>1088 ;
       } ;
drop seam* subsize ; save temp8, replace ;
use temp8 ; keep pid idate* ; sort pid ; save temp9, replace ;
use temp8 ; reshape long wave stat sdate edate, i(pid) j(spell) ;
drop if stat==. ;
sort pid ; merge pid using temp9 ; tab _merge ; drop _merge ;

end ;


************************** ReshapeB ****************************************;
/* ReshapeB: reshapes the data from wide to long for method B */

capture program drop ReshapeB ;
program define ReshapeB ;

egen size=max(total) ; local s=size ; save temp1, replace ;

use temp1 ; keep if total<10 ;
drop wave10-wave`s' stat10-stat`s' sdate10-sdate`s' edate10-edate`s' ;
save temp2, replace ; drop idate* ;
reshape long wave stat sdate edate, i(pid total) j(spell) ;
drop if stat==. ; sort pid ; save temp3, replace ;
use temp2 ; keep pid idate* ; sort pid ; merge pid using temp3 ;
tab _merge ; drop _merge ; save temp4, replace ;

use temp1 ; keep if total>=10 & total<16 ;
drop wave16-wave`s' stat16-stat`s' sdate16-sdate`s' edate16-edate`s' ;
save temp5, replace ; drop idate* ;
reshape long wave stat sdate edate, i(pid total) j(spell) ;
drop if stat==. ; sort pid ; save temp6, replace ;
use temp5 ; keep pid idate* ; sort pid ; merge pid using temp6 ;
tab _merge ; drop _merge ; save temp7, replace ;

use temp1 ; keep if total>=16 & total<25 ;
drop wave25-wave`s' stat25-stat`s' sdate25-sdate`s' edate25-edate`s' ;
save temp8, replace ; drop idate* ;
reshape long wave stat sdate edate, i(pid total) j(spell) ;
```

```
drop if stat==. ; sort pid ; save temp9, replace ;
use temp8 ; keep pid idate* ; sort pid ; merge pid using temp9 ;
tab _merge ; drop _merge ; save temp10, replace ;

use temp1 ; keep if total>=25 ; save temp11, replace ; drop idate* ;
reshape long wave stat sdate edate, i(pid total) j(spell) ;
drop if stat==. ; sort pid ; save temp12, replace ;
use temp11 ; keep pid idate* ; sort pid ; merge pid using temp12 ;
tab _merge ; drop _merge ; save temp13, replace ;

use temp4 ; append using temp7 ; append using temp10 ; append using temp13 ;
drop size ;

end ;


************************** Divide ******************************************;
/* Divide: divides the data into interviews containing only spells from that
   interview and prior interviews */

capture program drop Divide ;
program define Divide ;

save temp1, replace ;
use temp1 ; keep if wave<=1 ; rename idate1 intdate ; drop idate* total ;
keep if intdate~=. ; gen count=1 ; sort pid ;
quietly egen total=sum(count), by(pid) ; drop count ;
replace edate=intdate if spell==total ;
reshape wide wave stat sdate edate, i(pid total intdate) j(spell) ;
gen intwave=1 ; save temp2, replace ;
use temp1 ; keep if wave<=2 ; rename idate2 intdate ; drop idate* total ;
keep if intdate~=. ; gen count=1 ; sort pid ;
quietly egen total=sum(count), by(pid) ; drop count ;
replace edate=intdate if spell==total ;
reshape wide wave stat sdate edate, i(pid total intdate) j(spell) ;
gen intwave=2 ; save temp3, replace ;
use temp1 ; keep if wave<=3 ; rename idate3 intdate ; drop idate* total ;
keep if intdate~=. ; gen count=1 ; sort pid ;
quietly egen total=sum(count), by(pid) ; drop count ;
replace edate=intdate if spell==total ;
reshape wide wave stat sdate edate, i(pid total intdate) j(spell) ;
gen intwave=3 ; save temp4, replace ;
use temp1 ; keep if wave<=4 ; rename idate4 intdate ; drop idate* total ;
keep if intdate~=. ; gen count=1 ; sort pid ;
quietly egen total=sum(count), by(pid) ; drop count ;
replace edate=intdate if spell==total ;
reshape wide wave stat sdate edate, i(pid total intdate) j(spell) ;
gen intwave=4 ; save temp5, replace ;
use temp1 ; keep if wave<=5 ; rename idate5 intdate ; drop idate* total ;
keep if intdate~=. ; gen count=1 ; sort pid ;
quietly egen total=sum(count), by(pid) ; drop count ;
replace edate=intdate if spell==total ;
reshape wide wave stat sdate edate, i(pid total intdate) j(spell) ;
gen intwave=5 ; save temp6, replace ;
use temp1 ; keep if wave<=6 ; rename idate6 intdate ; drop idate* total ;
keep if intdate~=. ; gen count=1 ; sort pid ;
quietly egen total=sum(count), by(pid) ; drop count ;
replace edate=intdate if spell==total ;
reshape wide wave stat sdate edate, i(pid total intdate) j(spell) ;
gen intwave=6 ; save temp7, replace ;
use temp1 ; keep if wave<=7 ; rename idate7 intdate ; drop idate* total ;
keep if intdate~=. ; gen count=1 ; sort pid ;
quietly egen total=sum(count), by(pid) ; drop count ;
replace edate=intdate if spell==total ;
reshape wide wave stat sdate edate, i(pid total intdate) j(spell) ;
gen intwave=7 ; save temp8, replace ;
use temp1 ; keep if wave<=8 ; rename idate8 intdate ; drop idate* total ;
keep if intdate~=. ; gen count=1 ; sort pid ;
quietly egen total=sum(count), by(pid) ; drop count ;
replace edate=intdate if spell==total ;
reshape wide wave stat sdate edate, i(pid total intdate) j(spell) ;
gen intwave=8 ; save temp9, replace ;
use temp1 ; keep if wave<=9 ; rename idate9 intdate ; drop idate* total ;
keep if intdate~=. ; gen count=1 ; sort pid ;
quietly egen total=sum(count), by(pid) ; drop count ;
replace edate=intdate if spell==total ;
```

```
reshape wide wave stat sdate edate, i(pid total intdate) j(spell) ;
gen intwave=9 ; save temp10, replace ;
use temp1 ; keep if wave<=10 ; rename idate10 intdate ; drop idate* total ;
keep if intdate~=. ; gen count=1 ; sort pid ;
quietly egen total=sum(count), by(pid) ; drop count ;
replace edate=intdate if spell==total ;
reshape wide wave stat sdate edate, i(pid total intdate) j(spell) ;
gen intwave=10 ; save temp11, replace ;
use temp1 ; keep if wave<=11 ; rename idate11 intdate ; drop idate* total ;
keep if intdate~=. ; gen count=1 ; sort pid ;
quietly egen total=sum(count), by(pid) ; drop count ;
replace edate=intdate if spell==total ;
reshape wide wave stat sdate edate, i(pid total intdate) j(spell) ;
gen intwave=11 ; save temp12, replace ;

use temp2 ; append using temp3 ; append using temp4 ; append using temp5 ;
append using temp6 ; append using temp7 ; append using temp8 ;
append using temp9 ; append using temp10 ; append using temp11 ;
append using temp12 ;
gen temp=intwave ; drop intwave ; rename temp intwave ;

end ;


************************** AmendC ****************************************;
/* AmendC: selects the observations and divides the data for SubAmend */

capture program drop AmendC ;
program define AmendC ;

sort pid intwave; save temp1, replace ;

use addinfo ; keep pid wave left idate ; save temp5, replace ;
use temp5 ; keep pid wave idate ; reshape wide idate, i(pid) j(wave) ;
sort pid ; save temp6, replace ;
use temp5 ; keep pid wave left ; sort pid ; merge pid using temp6 ;
tab _merge ; drop _merge ;
rename wave intwave ; sort pid intwave ; merge pid intwave using temp1 ;
tab _merge ; keep if _merge==3 ; drop _merge ;
egen size=max(total) ; save temp7, replace ;

local s=size ;
use temp7 ; keep if total<4 ; gen subsize=3 ;
drop wave4-wave`s' stat4-stat`s' sdate4-sdate`s' edate4-edate`s' ;
SubAmend ; save temp8, replace ;
use temp7 ; keep if total>=4 & total<7 ; gen subsize=6 ;
drop wave7-wave`s' stat7-stat`s' sdate7-sdate`s' edate7-edate`s' ;
SubAmend ; save temp9, replace ;
use temp7 ; keep if total>=7 & total<11 ; gen subsize=10 ;
drop wave11-wave`s' stat11-stat`s' sdate11-sdate`s' edate11-edate`s' ;
SubAmend ; save temp10, replace ;
use temp7 ; keep if total>=11 & total<16 ; gen subsize=15 ;
drop wave16-wave`s' stat16-stat`s' sdate16-sdate`s' edate16-edate`s' ;
SubAmend ; save temp11, replace ;
use temp7 ; keep if total>=16 & total<25 ; gen subsize=24 ;
drop wave25-wave`s' stat25-stat`s' sdate25-sdate`s' edate25-edate`s' ;
SubAmend ; save temp12, replace ;
use temp7 ; keep if total>=25 ; egen subsize=max(total) ;
SubAmend ; save temp13, replace ;

use temp8 ; append using temp9 ; append using temp10 ; append using temp11 ;
append using temp12 ; append using temp13 ;
drop size ;

end ;


************************** SubAmend ****************************************;
/* SubAmend: amends the data for method C */

capture program drop SubAmend ;
program define SubAmend ;

replace sdate1=left if sdate1==. & (edate1-left<12) & (edate1-left>=0) ;

/* loop around spells 1 to subsize */
```

```
local i=0 ; local j=0 ;
while `i' <= (subsize-2) { ;
        local i = `i'+1 ; local j = `i'+1 ;
        quietly gen match`i'=1 if stat`j'==99 & edate`j'==. &
                (sdate`i'-sdate`j'<3) & (sdate`i'-sdate`j'>-3) ;
        quietly replace stat`j'=stat`i' if match`i'==1 ;
        quietly replace edate`j'=edate`i' if match`i'==1 ;
        quietly gen first`j'=0 if wave`j'==0 ;
        quietly replace first`j'=1088 if wave`j'==1 ;
        quietly replace first`j'=idate1 if wave`j'==2 ;
        quietly replace first`j'=idate2 if wave`j'==3 ;
        quietly replace first`j'=idate3 if wave`j'==4 ;
        quietly replace first`j'=idate4 if wave`j'==5 ;
        quietly replace first`j'=idate5 if wave`j'==6 ;
        quietly replace first`j'=idate6 if wave`j'==7 ;
        quietly replace first`j'=idate7 if wave`j'==8 ;
        quietly replace first`j'=idate8 if wave`j'==9 ;
        quietly gen last`j'=1088 if wave`j'==0 ;
        quietly replace last`j'=idate1 if wave`j'==1 ;
        quietly replace last`j'=idate2 if wave`j'==2 ;
        quietly replace last`j'=idate3 if wave`j'==3 ;
        quietly replace last`j'=idate4 if wave`j'==4 ;
        quietly replace last`j'=idate5 if wave`j'==5 ;
        quietly replace last`j'=idate6 if wave`j'==6 ;
        quietly replace last`j'=idate7 if wave`j'==7 ;
        quietly replace last`j'=idate8 if wave`j'==8 ;
        quietly replace last`j'=idate9 if wave`j'==9 ;
        quietly gen new`i'=1 if edate`i'==. & sdate`j'==.
                        & sdate`i'~=. & edate`j'~=. ;
        quietly replace edate`i'=
                        (max(sdate`i',first`j')+min(edate`j',last`j')) / 2
                        if new`i'==1 ;
        quietly replace sdate`j'=edate`i' if new`i'==1 ;
        quietly gen before`j'=1088 if wave`j'==1 ;
        quietly replace before`j'=idate1 if wave`j'==2 ;
        quietly replace before`j'=idate2 if wave`j'==3 ;
        quietly replace before`j'=idate3 if wave`j'==4 ;
        quietly replace before`j'=idate4 if wave`j'==5 ;
        quietly replace before`j'=idate5 if wave`j'==6 ;
        quietly replace before`j'=idate6 if wave`j'==7 ;
        quietly replace before`j'=idate7 if wave`j'==8 ;
        quietly replace before`j'=idate8 if wave`j'==9 ;
        quietly replace edate`i'=sdate`j' if ((sdate`j'-edate`i')>0)
                        & sdate`j'~=. & edate`i'~=. &
                        (before`j'~=. | (wave`i'==wave`j')) ;
        } ;
drop idate* new* first* last* before* ;

gen educ=0 ;
/* loop around spells 1 to subsize */
local i=0 ; local j=0 ;
while `i' <= (subsize-2) { ;
        local i = `i'+1 ; local j = `i'+1 ;
        quietly replace educ=1 if stat`i'==7 ;
        quietly gen omit`i'=1 if stat`i'==99 & edate`i'>sdate`j'
                        & edate`i'~=. ;
        quietly replace omit`i'=1 if stat`i'==99 &
                        (((sdate`i'-sdate`j'<3) & (sdate`i'-sdate`j'>-3)) |
                        ((edate`i'-edate`j'<3) & (edate`i'-edate`j'>-3))) ;
        quietly replace omit`i'=2 if stat`j'==99 & edate`i'>sdate`j'
                        & edate`i'~=. ;
        quietly replace omit`i'=2 if match`i'==1 ;
        quietly replace omit`i'=2 if stat`j'==99 & (sdate`i'-sdate`j'<3)
                        & (sdate`i'-sdate`j'>-3) ;
        quietly replace omit`i'=3 if edate`i'==. & sdate`i'>=sdate`j'
                        & sdate`i'~=. & stat`i'==stat`j' ;
        quietly replace omit`i'=4 if edate`i'==. & stat`i'==stat`j'
                        & stat`i'~=2 ;
        quietly replace omit`i'=5 if sdate`j'==. & (edate`j'-edate`i'<=12)
                        & (stat`i'==stat`j') ;
        quietly replace omit`i'=6 if edate`i'==edate`j' & stat`i'==stat`j'
                        & (sdate`i'==. | sdate`j'==.) ;
        quietly replace sdate`j'=sdate`i' if omit`i'>0 & omit`i'<6
                        & ~(omit`i'==2 & sdate`i'==.) ;
        quietly replace sdate`j'=sdate`i' if omit`i'==6 & sdate`i'~=. ;
        quietly replace stat`j'=stat`i' if omit`i'==2 ;
```

```
        quietly replace sdate`j'=edate`i' if sdate`j'==.
                        & (edate`j'-edate`i'<=12) & (stat`i'~=stat`j') ;
        quietly replace wave`j'=wave`i' if omit`i'>0 & omit`i'~=. ;
        } ;
drop match* subsize ;
reshape long stat sdate edate wave omit,
        i(pid total intwave intdate left educ) j(spell) ;
drop if stat==. ; drop if omit>0 & omit~=. ; drop omit ;
gen count=1 ; drop total ; sort pid intwave ;
egen total=sum(count), by(pid intwave) ;
drop if spell==1 & stat==7 & total>1 ;
rename spell temp ; sort pid intwave temp ;
quietly by pid intwave: gen spell=sum(count) ; drop temp total ;
sort pid intwave ; egen total=sum(count), by(pid intwave) ;
drop if spell==1 & stat==7 & total>1 ; rename spell temp ;
sort pid intwave temp ; quietly by pid intwave: gen spell=sum(count) ;
drop temp total ; sort pid intwave ; egen total=sum(count), by(pid intwave) ;
drop if spell==1 & stat==7 & total>1 ;
rename spell temp ; sort pid intwave temp ;
quietly by pid intwave: gen spell=sum(count) ;
drop total temp ; sort pid ; egen total=sum(count), by(pid intwave) ;
drop count ;
reshape wide stat sdate edate wave,
        i(pid total intwave intdate left educ) j(spell) ;
egen smin=rmin(sdate*) ; replace left=smin if left<smin & smin~=. & educ==1 ;
drop smin educ ;

/* do this loop 10 times */
local t=0 ;
while `t'<10 { ;
        local t = `t'+1 ;
        egen subsize=max(total) ;
        gen omit1=0 ;
        /* loop around spells 1 to subsize */
        local i=0 ; local j=0 ;
        while `i' <= (subsize-2) { ;
                local i = `i'+1 ; local j = `i'+1 ;
                quietly gen omit`j'=1 if sdate`j'<edate`i' & edate`i'==edate`j'
                        & edate`i'~=. & ~(wave`i'==0 & wave`j'==1) ;
                quietly replace sdate`j'=edate`i' if sdate`j'<edate`i'
                        & edate`i'~=. & ~(wave`i'==0 & wave`j'==1) ;
                quietly replace omit`i'=1 if sdate`j'<=sdate`i' & edate`i'<=edate`j'
                        & edate`j'~=. & (wave`i'==0 & wave`j'==1) ;
                quietly replace wave`j'=wave`i' if omit`i'==1 ;
                quietly replace edate`i'=sdate`j' if sdate`j'<edate`i' & edate`i'~=.
                        & (wave`i'==0 & wave`j'==1) ;
                } ;
        reshape long stat sdate edate wave omit,
                i(pid total intwave intdate left) j(spell) ;
        drop if stat==. ; drop if omit>0 & omit~=. ; drop omit subsize ;
        rename spell temp ; gen count=1 ; sort pid intwave temp ;
        quietly by pid intwave: gen spell=sum(count) ;      drop total temp ;
        sort pid ; egen total=sum(count), by(pid intwave) ; drop count ;
        reshape wide stat sdate edate wave,
                i(pid total intwave intdate left) j(spell) ;
        } ;
replace sdate1=left if sdate1==. & (edate1-left<12) & (edate1-left>=0) ;
rename left mendleft ;

egen subsize=max(total) ;
/* loop around spells 1 to subsize */
local i=0 ; local j=0 ;
while `i' <= (subsize-2) { ;
        local i = `i'+1 ; local j = `i'+1 ;
        quietly gen omit`i'=1 if stat`i'==stat`j' & stat`i'~=2
                & edate`i'>=sdate`j' & edate`i'~=. ;
        quietly replace omit`i'=1 if stat`i'==stat`j' & stat`i'==2
                & edate`i'>=sdate`j' & edate`i'~=. & wave`i'==0 & wave`j'==1 ;
        quietly replace sdate`j'=sdate`i' if omit`i'==1 ;
        quietly replace wave`j'=wave`i' if omit`i'==1 ;
        } ;
reshape long stat sdate edate wave omit,
        i(pid total intwave intdate mendleft) j(spell) ;
drop if stat==. ; drop if omit==1 ; drop omit subsize ;
rename spell temp ; gen count=1 ; sort pid intwave temp ;
quietly by pid intwave: gen spell=sum(count) ;
```

```
drop total temp ; sort pid ; egen total=sum(count), by(pid intwave) ; drop count ;
reshape wide stat sdate edate wave, i(pid total intwave intdate mendleft) j(spell) ;

end ;


************************** SelectB ****************************************;
/* SelectB: divides the data for SubSB */

capture program drop SelectB ;
program define SelectB ;

egen size=max(total) ; save temp1, replace ;
local s=size ;
use temp1 ; keep if total<4 ; gen subsize=3 ;
drop wave4-wave`s' stat4-stat`s' sdate4-sdate`s' edate4-edate`s' ;
SubSB ; save temp2, replace ;
use temp1 ; keep if total>=4 & total<7 ; gen subsize=6 ;
drop wave7-wave`s' stat7-stat`s' sdate7-sdate`s' edate7-edate`s' ;
SubSB ; save temp3, replace ;
use temp1 ; keep if total>=7 & total<11 ; gen subsize=10 ;
drop wave11-wave`s' stat11-stat`s' sdate11-sdate`s' edate11-edate`s' ;
SubSB ; save temp4, replace ;
use temp1 ; keep if total>=11 & total<16 ; gen subsize=15 ;
drop wave16-wave`s' stat16-stat`s' sdate16-sdate`s' edate16-edate`s' ;
SubSB ; save temp5, replace ;
use temp1 ; keep if total>=16 & total<25 ; gen subsize=24 ;
drop wave25-wave`s' stat25-stat`s' sdate25-sdate`s' edate25-edate`s' ;
SubSB ; save temp6, replace ;
use temp1 ; keep if total>=25 ; egen subsize=max(total) ;
SubSB ; save temp7, replace ;
use temp2 ; append using temp3 ; append using temp4 ; append using temp5 ;
append using temp6 ; append using temp7 ; drop size ;

end ;


************************** SubSB ****************************************;
/* SubSB: selects spells for method B by dropping spells which are included in a more
recent interview */

capture program drop SubSB ;
program define SubSB ;

gen cut=9999 ;
/* loop around spells from subsize down to 1 */
local i=subsize+1 ;
while `i' > 1 { ;
        local i = `i'-1 ;
        quietly gen omit`i'=0 ;
        quietly replace omit`i'=1 if sdate`i'>=cut & sdate`i'~=. ;
        quietly replace cut=sdate`i' if omit`i'==0 & sdate`i'<cut ;
        } ;
drop cut subsize total ;
reshape long wave stat sdate edate omit, i(pid intdate intwave) j(spell) ;
drop if stat==. ; drop if omit==1 ; drop omit ; rename spell temp ;
gen count=1 ; sort pid intwave temp ;
quietly by pid intwave: gen spell=sum(count) ;
quietly egen total=sum(count), by(pid intwave) ; drop temp count ;
reshape wide wave stat sdate edate, i(pid total intdate intwave) j(spell) ;
/* loop around spells 1 to subsize */
egen subsize=max(total) ;
local i=0 ; local j=0 ;
while `i' <= subsize-2 { ;
        local i = `i'+1 ; local j = `i'+1 ;
        quietly gen omit`i'=0 ;
        quietly replace omit`i'=1 if stat`i'==stat`j' & stat`i'~=2
                        & (sdate`j'-edate`i'<1) ;
        quietly replace omit`i'=1 if stat`i'==stat`j' & stat`i'==2
                        & (edate`i'>sdate`j') & edate`i'~=. ;
        quietly replace sdate`j'=sdate`i' if omit`i'==1 ;
        quietly replace wave`j'=wave`i' if omit`i'==1 ;
        quietly replace edate`i'=sdate`j' if edate`i'>sdate`j' ;
        } ;
drop subsize ;
reshape long wave stat sdate edate omit, i(pid  total intdate intwave) j(spell) ;
```

```
drop if stat==. ; drop if omit==1 ; drop total omit ; rename spell temp ;
gen count=1 ; sort pid intwave temp ;
quietly by pid intwave: gen spell=sum(count) ;
quietly egen total=sum(count), by(pid intwave) ; drop temp count ;
reshape wide wave stat sdate edate, i(pid total intdate intwave) j(spell) ;

end ;


************************** Problems ******************************************;
/* Problems: checks problems */

capture program drop Problems ;
program define Problems ;

/* spell problems */

egen size=max(total) ;
/* first spell */
quietly gen prob1=0 ;
quietly replace prob1=prob1+2 if stat1==99 ;
quietly replace prob1=prob1+3 if stat2==99 ;
quietly replace prob1=prob1+20 if sdate1==. & stat1~=. ;
quietly replace prob1=prob1+200 if edate1==. & stat1~=. ;
quietly replace prob1=prob1+2000 if sdate1>edate1 & sdate1~=. ;
quietly replace prob1=prob1+30000 if edate1>sdate2 & edate1~=. ;
quietly replace prob1=prob1+300000 if sdate2-edate1>1 & sdate2~=.
               & edate1~=. & ~(wave2==0 & datasource==2) ;
/* loop around spells 2 to size-1 */
local i=0 ; local j=1 ; local k=0 ;
while `j' <= (size-2) { ;
       local j = `j'+1 ; local i = `j'-1 ; local k = `j'+1 ;
       quietly gen prob`j'=0 ;
       quietly replace prob`j'=prob`j'+1 if stat`i'==99 ;
       quietly replace prob`j'=prob`j'+2 if stat`j'==99 ;
       quietly replace prob`j'=prob`j'+3 if stat`k'==99 ;
       quietly replace prob`j'=prob`j'+20 if sdate`j'==. & stat`j'~=. ;
       quietly replace prob`j'=prob`j'+200 if edate`j'==. & stat`j'~=. ;
       quietly replace prob`j'=prob`j'+2000 if sdate`j'>edate`j'
              & sdate`j'~=. ;
       quietly replace prob`j'=prob`j'+10000 if edate`i'>sdate`j'
              & edate`i'~=. ;
       quietly replace prob`j'=prob`j'+30000 if edate`j'>sdate`k'
              & edate`j'~=. ;
       quietly replace prob`j'=prob`j'+100000 if sdate`j'-edate`i'>1
              & sdate`j'~=. & edate`i'~=. & ~(wave`i'==0 & datasource==2) ;
       quietly replace prob`j'=prob`j'+300000 if sdate`k'-edate`j'>1
              & sdate`k'~=. & edate`j'~=. & ~(wave`j'==0 & datasource==2) ;
       } ;
/* last spell */
local i=size-1 ; local j=size ;
quietly gen prob`j'=0 ;
quietly replace prob`j'=prob`j'+1 if stat`i'==99 ;
quietly replace prob`j'=prob`j'+2 if stat`j'==99 ;
quietly replace prob`j'=prob`j'+20 if sdate`j'==. & stat`j'~=. ;
quietly replace prob`j'=prob`j'+200 if edate`j'==. & stat`j'~=. ;
quietly replace prob`j'=prob`j'+2000 if sdate`j'>edate`j' & sdate`j'~=. ;
quietly replace prob`j'=prob`j'+10000 if edate`i'>sdate`j' & edate`i'~=. ;
quietly replace prob`j'=prob`j'+100000 if sdate`j'-edate`i'>1 & sdate`j'~=.
                     & edate`i'~=. & ~(wave`i'==0 & datasource==2) ;

/* history summary problems */

gen tprob=0 ; gen eprob=0 ; gen mprob=0 ; gen sprob=0 ;
/* loop around spells 1 to size-1 */
local i=0 ; local j=0 ;
while `i' <= (size-2) { ;
       local i = `i'+1 ; local j = `i'+1 ;
       quietly replace tprob=tprob+1 if prob`i'~=0 & `i'==total ;
       quietly replace tprob=tprob+100 if sdate`i'<1088 & `i'==total
                     & datasource==3 ;
       quietly replace eprob=eprob+1 if (stat`i'==1 | stat`i'==2) & prob`i'>0 ;
       quietly replace mprob=mprob+1 if stat`i'==99 ;
       quietly replace sprob=sdate`j' if prob`i'>0 & `i'<total ;
       quietly replace sprob=sdate`i' if prob`i'>0 & sdate`i'>sdate`j'
                     & sdate`i'~=. ;
```

```
        quietly replace sprob=edate`i' if prob`i'>0 & `i'==total ;
        } ;
/* last spell */
local j=size ;
quietly replace tprob=tprob+1 if prob`j'~=0 & `j'==total ;
quietly replace tprob=tprob+100 if sdate`j'<1088 & `j'==total & datasource==3 ;
quietly replace eprob=eprob+1 if (stat`j'==1 | stat`j'==2) & prob`j'>0 ;
quietly replace mprob=mprob+1 if stat`j'==99 ;
quietly replace sprob=intdate if prob`j'>0 ;
replace eprob=eprob+mprob ; drop mprob ;
sort pid intwave ; save temp1, replace ;
use addinfo ; keep pid wave left ; rename wave intwave ;
sort pid intwave ; merge pid intwave using temp1 ;
tab _merge ; keep if _merge==3 ; drop _merge ;
while method==3 { ;
        replace left=mendleft if mendleft~=. ;
        drop mendleft ;
        replace method=method+99 ;
        } ;
replace method=method-99 if method>99 ;
replace tprob=tprob+10 if total==1 & wave1~=0 & sdate1>left ;
replace eprob=eprob+100 if wave1~=0 & (((sdate1-left>3) & sdate1~=.) | left==.) ;

drop size ; egen smin=rmin(sdate*) ;

end ;


************************** ExperTen *****************************************;
/* ExperTen: adds the experience and employer tenure variables */

capture program drop ExperTen ;
program define ExperTen ;

egen size=max(total) ;
gen tenure=0 ; gen exper=0 ; gen currstat=0 ;
/* loop around spells 1 to size-1 */
local i=0 ; local j=0 ;
while `i' <= (size-2) { ;
        local i = `i'+1 ; local j = `i'+1 ;
        quietly gen len`i'=edate`i'-sdate`i' ;
        quietly replace len`i'=0.5 if len`i'==0 ;
        quietly replace currstat=stat`i' if `i'==total ;
        quietly replace tenure=len`i' if (stat`i'==1 | stat`i'==2)
                & `i'==total & len`i'~=. ;
        quietly replace exper=exper+len`i' if (stat`i'==1 | stat`i'==2)
                & len`i'~=. ;
        } ;
local k=size ;
quietly gen len`k'=edate`k'-sdate`k' ;
quietly replace len`k'=0.5 if len`k'==0 ;
quietly replace currstat=stat`k' if `k'==total ;
quietly replace tenure=len`k' if (stat`k'==1 | stat`k'==2) & `k'==total
                & len`k'~=. ;
quietly replace exper=exper+len`k' if (stat`k'==1 | stat`k'==2) & len`k'~=. ;
quietly replace tenure=. if tprob~=0 ; quietly replace exper=. if eprob~=0 ;
quietly replace tenure=. if currstat~=1 & currstat~=2 ;
quietly drop len* ;

label define stat 1 "self-employed" 2 "employed" 3 "unemployed"
                  4 "out of labour force" 7 "full-time education"
                  99 "missing" ;
/* loop around spells 1 to size */
local i=0 ;
while `i'<=(size-1) { ;
        local i=`i'+1 ;
        quietly label values stat`i' stat ;
        } ;
quietly label values currstat stat ;
quietly drop size ;

end ;
```

```
/*******************************************************************************
                                RUNNING THE SHOW
*******************************************************************************/

use bcpanel ; drop sflag* eflag* ; CrosWave ; save tempema1, replace ;
SeamsA ; save tempema2, replace ; Divide ; save tempema3, replace ;
gen method=1 ; gen datasource=1 ; Problems ; ExperTen ; save histA1, replace ;

use bcpanel ; drop sflag* eflag* ; ReshapeB ; save tempemb1, replace ;
Divide ; save tempemb2, replace ; SelectB ; save tempemb3, replace ;
gen method=2 ; gen datasource=1 ; Problems ; ExperTen ; save histB1, replace ;

use bcpanel ; DateFixEM ; save tempemc1, replace ; CrosWave ; save tempemc2, replace ;
Divide ; save tempemc3, replace ; AmendC ; save tempemc4, replace ;
gen method=3 ; gen datasource=1 ; Problems ; ExperTen ; save histC1, replace ;

use bcpanel ; drop sflag* eflag* ; CrosWave ; save tempemd1, replace ;
Divide ; save tempemd2, replace ;
gen method=4 ; gen datasource=1 ; Problems ; ExperTen ; save histD1, replace ;

use cpanel ; drop sflag* eflag* ; CrosWave ; save tempwa1, replace ;
SeamsA ; save tempwa2, replace ;  Divide ; save tempwa3, replace ;
gen method=1 ; gen datasource=2 ; Problems ; ExperTen ; save histA2, replace ;

use cpanel ; drop sflag* eflag* ; ReshapeB ; save tempwb1, replace ;
Divide ; save tempwb2, replace ; SelectB ; save tempwb3, replace ;
gen method=2 ; gen datasource=2 ; Problems ; ExperTen ; save histB2, replace ;

use cpanel ; DateFixW ; save tempwc1, replace ; CrosWave ; save tempwc2, replace ;
Divide ; save tempwc3, replace ; AmendC ; save tempwc4, replace ;
gen method=3 ; gen datasource=2 ; Problems ; ExperTen ; save histC2, replace ;

use cpanel ; drop sflag* eflag* ; CrosWave ; save tempwd1, replace ;
Divide ; save tempwd2, replace ;
gen method=4 ; gen datasource=2 ; Problems ; ExperTen ; save histD2, replace ;

log close ;
set more 0 ;
#delimit cr
```

BIASES IN THE REPORTING OF LABOUR MARKET DYNAMICS

Gillian Paull

THE INSTITUTE FOR FISCAL STUDIES
WP02/10

Biases in the Reporting of Labour Market Dynamics

Gillian Paull[§]

The Institute for Fiscal Studies

June 2002

JEL classification: J60 J64 C8
Keywords: labour market dynamics, recall bias

Acknowledgements

This paper was written with financial support from the Leverhulme Trust (while the author was at the Centre for Economic Performance, London School of Economics) and from the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at the Institute of Fiscal Studies. It is a revised version of Paull (1997). The author wishes to thank Paul Gregg, David Card, Anne Case, Cecilia Rouse, Shirley Dex, Alison Booth and participants at a CEP seminar and University of Essex seminar for useful comments. The data used in this paper were made available through the ESRC Data Archive. The data were originally collected by the ESRC Research Centre on Micro-social Change at the University of Essex (now incorporated within the Institute for Social and Economic Research). Neither the original collectors of the data nor the Archive bear any responsibility for the analyses or interpretations presented here.

[§] Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE, UK. Tel: 020 7291 4842. E-mail: gill_p@ifs.org.uk

Executive Summary

Correctly measuring individual dynamics in labour market behaviour has become increasingly important as research and policy attention has become more focused on the relationships between current employment opportunities and past experience. Surveys collecting information on labour market histories use repeated interviews and retrospective reporting, laying the resulting data open to potential biases from spurious transitions due to random measurement errors and from systematic recall error. This paper uses a unique data opportunity provided by the British Household Panel Survey to systematically investigate the impact of recall on measured labour market behaviour and to highlight how and to what degree the biases in the reported data may affect the estimation of models of labour market dynamics. The results allow analysts to judge whether conclusions drawn from such models are likely to be compromised by the reporting biases.

1. Introduction

Academic debate and policy discussion involving the labour market has increasingly focused on the *dynamics* of behaviour, emphasising the importance of changes in work choices and the relationships between past patterns of employment and current options. For example, recent research has examined the effect of unemployment on future employment and earnings[1]; the impact of employment experience and job tenure on wages[2]; and the role of career profiles in explaining gender wage differentials[3]. On the policy front, programs designed to minimize the duration of unemployment (such as the New Deal programs) and financial incentives for the low paid to remain in employment (such as the Working Families' Tax Credit) have been at the forefront of UK initiatives, while welfare policy in the US has increasingly emphasised the role of time limits for state support for single mothers. In analysing these types of issues, it is essential to have an accurate picture of dynamics within the labour market, but obtaining an unbiased picture from survey data may not be easy[4].

One approach to collecting information on labour market dynamics is to repeatedly ask individuals for their current labour market status, as would be collected in the repeated interviews of a panel survey. Ideally, the interval of time between the data points would be as short as possible to ensure that all changes in state are captured, but this may be limited by practical considerations[5] or could exacerbate some of the problems of panel surveys[6]. Moreover, Poterba & Summers (1986) have shown that transition rates calculated from this type of data may overstate the degree of dynamics in the labour market because classification errors in the reported labour market status can generate spurious transitions between states. Indeed, their estimates suggest that reporting errors lead to a considerable understatement of the duration of

[1] For example, see Machin & Manning (1999), Arulampalam, Booth & Taylor (2000), Arulampalam, Gregg & Gregory (2001), Arulampalam (2001), Gregg (2001) and Gregory & Jukes (2001), for evidence from the UK. See Stevens (1997) and Kletzer (1998) for the US.

[2] For example, see Booth & Frank (1995) and Manning (1998b) for studies from the UK; Altonji & Williams (1998) and Bronars & Famulari (1997) for the US; and Dustmann & Meghir (2001) for Germany and the US.

[3] For example, see Harkness (1996), Manning (1998a), Manning & Robinson (1998) and Myck & Paull (2001) for research on the UK. See Bowlus (1997) and Light & Ureta (1995) for the US. Gender differences in the UK labour market are also examined in Booth, Francesconi & Garcia-Serrano (1999), Booth, Jenkins & Garcia Serrano (1999) and Booth & Francesconi (2000).

[4] Administrative records may provide an alternative source of data, such as, the matched NESPD-JUVOS dataset described in Gregory & Jukes (2001). However, such administrative data rarely provides as detailed household information as surveys and tends to only measure employment spells and benefit-related unemployment spells, omitting to record time out of the labour force directly.

[5] One possibility would be a self-completed diary of labour market status, but even diary records are not without potential measurement error (see Dex (1995), page 65).

[6] There are several potential drawbacks of panel data. First, non-random attrition from the panel may generate biased model estimates (for example, see Peracchi & Welch (1995) or Paull (1997)). Second, there may be "time-in-sample" bias (sometimes called "panel conditioning"), where estimates from people with different levels of exposure to the survey have different expected response values. The combination of these two effects has been referred to as "rotation group bias" in the context of the CPS. Finally, there may be large financial and organisational investments involved in initiating and continuing the collection of panel data. A review of these issues can be found in Kalton, Kasprzyk & McMillen (1989).

unemployment and an overstatement of the frequency of labour force entry and exit[7]. It has been argued that the arbitrariness of the distinction between unemployment and out of the labour force may be particularly likely to generate spurious movements in and out of the labour force[8]. Relatedly, the "seam problem" (a tendency for reported changes in status to bunch in the period immediately after an interview) sometimes observed in retrospective data collected in a panel fashion may also be a consequence of measurement error in the reported current status[9].

An alternative approach to obtain the desired data is to ask individuals to retrospectively recall their behaviour over a specified prior period, either by requesting the dates of changes in behaviour or by asking for the main activity during a number of sub-periods[10]. However, the act of recollection may generate "recall" biases, whereby reported behaviour is not only subject to random errors but also systematic errors that may intensify as the period of recall increases. The evidence on the nature and importance of these potential recall biases is patchy, most having tended to focus only on the impact of recall error on unemployment. Several studies have shown that the reported aggregate time in unemployment falls with the length of recall[11] or that annual weeks of work increase as the recall period lengthens[12]. There is some limited evidence that shorter spells of unemployment are less likely to be recalled at a later date than longer spells[13] and that the reported length of unemployment spells may increase with recall[14]. The "seam problem" described above may also be an indication of recall error: the tendency for reported changes to bunch in the period immediately after an interview suggests that a change that may have occurred between interviews has simply been forgotten in the latter report.

The purpose of this paper is twofold. First, to provide a systematic investigation of the impact of recall on measured labour market dynamics when data is collected

---

[7] Poterba & Summers (1986) use US data from the Current Population Survey Reinterview Survey to estimate the incidence of errors in the gross changes data and develop a method for adjusting the measured transitions rates for spurious transitions.

[8] Clark & Summers (1979) argue that the way the data is collected in the US Current Population Survey generates an ambiguity and arbitrariness in the distinction between unemployment and out of the labour force, making it likely that "some observed entry and exit flows arise from inconsistent reporting of consistent behaviour" (page 28). Poterba & Summers (1984) also argue that the state of "not in the labor force" is functionally equivalent to unemployment (page 41). However, other work has concluded that unemployment and inactivity are distinct states (Flinn & Heckman (1983) and Gonul (1992)).

[9] For example, Hujer & Schneider (1989) show that the transition rate from unemployment to employment for men in the German Socio Economic Panel is significantly higher in December to January than in any other month. This data is collected by retrospectively asking individuals to report their status in each month of the prior calendar year and the years are then combined.

[10] Retrospective data is collected in several major cross-sectional surveys: the British Family and Working Lives Survey (asks about events since age 16); the lifetime histories in waves B and C of the British Household Panel Survey (covering time since leaving full-time education); and the March Work Experience Survey of the US Current Population Survey (requesting information from the previous calendar year). In addition, most of the major labour market *panel surveys* rely to some degree on *recalled* data for collecting information on labour market dynamics by asking respondents to retrospectively fill-in the gaps between interviews. In Britain, the Quarterly Labour Force Survey asks about the prior three months, the British Household Panel Survey collects information for the previous year and the National Child Development Survey covers the 7-10 years between interviews. In the US, the Panel Study of Income Dynamics requests information from the previous year and the Survey of Income and Program Participation looks back over four months. The Canadian Labour Market Activity Survey collects information on the prior calendar year (see Jones & Riddell (1995) for a description of this last survey).

[11] See Morgenstern & Barrett (1974), Horvath (1982), Akerlof & Yellen (1985), Duncan & Hill (1985), Mathiowetz & Duncan (1988), Levine (1993), Elias (1997) and Dex & McCulloch (1998).

[12] See Powers et al (1978) and Ryscavage & Coder (1989).

[13] See Mathiowetz & Duncan (1988) and Levine (1993).

[14] See Bowers & Horvath (1984) and Poterba & Summers (1984)

retrospectively. Second, to highlight how and to what degree the two biases (spurious transition and recall) may affect the estimation of models of labour market dynamics. Particular attention is paid to the accuracy of the reporting of behaviour across different types of respondent[15].

In considering the nature of recall bias, the analysis covers the complete labour market history in terms of a comprehensive sequence of spells. Although it is known that the reported time in unemployment tends to diminish as the length of recall period increases, the corresponding impacts on other states have not been established. Moreover, previous work has not examined the impact of recall on reported *transitions* between *all* types of labour market state. For example, are shorter spells of all types less likely to be recalled correctly than longer spells ? Are forgotten spells of unemployment redefined as a different state or are they subsumed into other spells ? This paper uses a unique data opportunity provided by the British Household Panel Survey (BHPS) to address these types of questions. By generating a period of overlap between the waves, the survey permits an examination of how reported labour market spells change when the same information is requested a year later. In particular, the data does not just measure changes in state at a particular point in time, but presents changes in the dynamic picture from an interval of time.

It is important to consider how and to what degree spurious transition bias and recall bias may affect the estimation of models of labour market dynamics. In particular, the biases may impact on aggregate statistics, but this does not automatically imply that the marginal properties and relationships with other variables will be affected in a significant manner[16]. As a means to test this, different methods (corresponding to alternative survey approaches) were used to construct individual labour market histories from the BHPS that would allow the potential impact of each bias on the measured behaviour to vary. Several models of the labour market were then estimated using the different construction methods: transition rates between labour market states; spell survival models; and the wage returns to experience and tenure. As well as generating substantive conclusions about labour market dynamics, the results calibrate how much measurement and recall error can influence these conclusions.

The remainder of the paper is organized as follows. The data source is briefly described in the next section. The third section presents the analysis of recall bias

---

[15] Previous work has indicated that the nature of the biases may differ by the gender and age of respondent. For example, see Morgenstern & Barrett (1974), Akerlof & Yellen (1985), Poterba & Summers (1986), Levine (1993), Poterba & Summers (1995) or Elias (1997).

[16] Holt, MacDonald & Skinner (1991) provide a framework for considering how recall error can affect methods of event history analysis. Several examples of how measurement errors can impact on estimates of labour market dynamics have previously been analysed. Poterba & Summers (1995) show that correcting for the classification errors strengthens the apparent effect of Unemployment Insurance on unemployment spell durations using US CPS data. Evidence from the US PSID in Brown & Light (1992) indicates that using inconsistent job tenure sequences can lead to misleading conclusions about the slope of wage-tenure profiles. Administrative and survey data from a large manufacturing company in the US suggest that survey data underestimates the return to tenure, but this is due to measurement error in the earnings variable rather than the tenure variable (Duncan & Hill (1985)).

using the periods of wave overlaps, beginning with a discussion of several hypotheses on the effects of recall. The following section considers the different methods for constructing individual labour market histories and examines the impact of the choice of method in estimating dynamic models of behaviour. The final section concludes.

2. The Data

The British Household Panel Survey (BHPS) is an annual survey of approximately 10,000 adults from a nationally representative sample of over 5,000 households. Individuals are re-interviewed in successive waves, together with any new adults living in the household. The vast majority of interviews are conducted during the autumn of each year, beginning in 1991 for the first wave, denoted wave A.

Information on jobs and periods of non-employment are collected in four sources[17]. First, at each interview, the individual is asked to select one of ten options best describing their current labour market state[18]. In addition, questions are asked whether the individual did any paid work or were away from a job in the week prior to interview[19]. The starting day/month/year for employment, self-employment and non-employment is recorded in 3 corresponding sets of variables[20]. For paid employment, this is the date of last promotion or employer change. Second, the survey also collects information about labour force behaviour between interviews. If the starting date for the current activity began on or after September 1 of the year prior to interview, the individual is asked what they did before the current state, selecting from 10 categories similar to those available for the current main activity[21], but corresponding information on whether the individual did any paid work or were away from a job is not collected for spells falling between interviews[22]. The starting date of any previous activity is also recorded[23], with employment spells again divided by whether there is an employer change or whether there is a change in position (promotion or job change) with the same employer. The pattern of questioning

[17] Further information on the collection of labour market history data in the BHPS can be found in Halpin (1997) and Oskrochi & Crouchley (2000).

[18] The question asking for current status is: "Please look at this card and tell me which best describes your current situation ?" providing the options: self employed, in paid employment (full or part time), unemployed, retired from paid work altogether, on maternity leave, looking after family or home, full-time student/at school, long term sick or disabled, on a government training scheme, something else. The responses are recorded in the variable wjbstat and, in waves other than wave A, in the variable wnemst

[19] The responses are recorded in the variables wjbhas and wjboff.

[20] Starting dates are recorded in the variables wjbbgd/m/y for those currently in paid employment, in jsbgd/m/y for those currently self-employed and in cjsbgd/m/y for those without a job.

[21] The question asking for prior states is: "Can you look at this card please and tell me which of the descriptions comes closest to what you were doing immediately before then ?" providing the options: doing a *different job* for the *same employer*, working for a *different employer*, in paid employment (not self employed), working for *myself* (self-employed), unemployed/looking for work, retired from paid work altogether, on maternity leave, looking after family or home, full-time education/student, *long term* sick or disabled, on a government training scheme, something else. The responses are recorded as the variable wjhstat.

[22] For those who report their main activity as work, the division between paid employment and self-employment (and between full-time and part-time employment) is recorded in the variable wjhsemp.

[23] The starting dates by day/month/year are recorded in the variables wjhbgd/m/y.

continues until the starting date of a spell is prior to the September 1 of the year prior to interview, covering the entire gap between interviews[24].

The third and fourth sources of information on labour market activity collect information on behaviour since the individual left full-time education. In the second wave (wave B), individuals are asked what they were doing after they left full-time education, selecting from 12 categories, again similar to those available for the current main activity[25]. They are then asked for the next date that the situation changed, recorded by month and year, and the category of the following activity[26]. This pattern of question is repeated until the current state is recorded. The dividing line between spells is change in main activity, with employment divided only into full/part time employment and into employee/self-employed rather than by employer change or promotion. Finally, in the third wave (wave C), information is collected on all jobs (other than the current employer) since leaving full-time education, divided into self-employment and employment[27]. Spells are divided by employer and both start and end dates, recorded by month and year, are collected because gaps of non-employment are not recorded as separate spells.

Hence, it appears that it should be relatively straightforward to compile a complete lifetime history consisting of spells of employment and non-employment, with the employment spells divided into different employers and/or self-employment and into part and full-time work and the non-employment spells divided into the 8 main categories. However, it should be noted that the information on labour market involvement is collected in two different ways: first by what individuals regard as there main activity ("main activity") and second, by whether they are involved in any type of employment ("any work"). The difference between these two could generate an inconsistent series of spells for individuals who hold a job, but regard their main activity as something other than employment[28]. In principle, it should be possible to distinguish between the two measures of labour market activity, generating two types of employment history: one consisting only of "main activity" and one consisting of all employment spells ("any work")[29]. However, there are two gaps in the survey design that mean that both histories may not be collected for all individuals. First, for the current spell at the time of interview, starting dates are collected for the main

[24] This second source (of between-wave information) is stored in the files wjobhist.

[25] The categories are: 1: self-employed, 2: full-time paid employment, 3: part-time paid employment, 4: unemployed, 5: retired from paid work altogether, 6: maternity leave, 7: looking after family or home, 8: full-time student / at school, 9: long-term sick or disabled, 10: on a government training scheme, 11: national service / war service, 12: something else. The responses are recorded in the variable bleshst.

[26] The starting dates by month and year are recorded in the variables bleshem/y.

[27] The job type (self-employed, full-time employee and part-time employee) is recorded in the variable cljsemp, while the starting months and years are recorded in the variables cljbgm/y and the end months and years are recorded in cljlftm/y.

[28] For example, spells of maternity leave or long-term sick may appear within a longer employment spell; temporary jobs held by students may appear within a longer education spell; and spells of retirement or family care may contain several different employment spells.

[29] Indeed, the collection of the two lifetime histories in the BHPS allows specifically for this distinction with the "main activity" collected in the second wave and the "any work" in the third wave.

activity only for those without any employment[30]. Second, spells between waves are collected only on the basis of the main activity, so that any employment not regarded as the main activity that falls completely between interviews is not recorded. In addition, as the current situation gives priority to jobs and the spells between interviews to main activity, inconsistencies are especially likely to arise when attempting to match spells across waves.

The prevalence of individuals who hold a job but regard their main activity as something other than employment is highlighted in table 1, which shows the distribution of the current labour force state reported in the first nine waves of the BHPS by gender and across three age groups[31]. The ten main activity categories are grouped into the five labour market states of employed, self-employed, unemployed, inactive and full-time education. Differences in behaviour are quite marked across the gender and the age groups. Men are more likely to be in paid employment than women and are much more likely to report being self-employed or unemployed than women. On the other hand, women are more likely to report an "inactive" category, usually family care, as their main activity than men, but the proportions of men and women in full-time education are roughly even. Not surprisingly, young people (under the age of 25) are much more likely to be in full-time education than older groups, while the eldest group (those over the age of 54) are most likely to be retired.

The proportions reporting that they did any paid work or were away from a job in the previous week are also shown in table 1 as those with "plus job". Most of those reporting their main activity as "maternity leave" or "government training" also held a job and are included in the paid employment total. Small fractions of the "unemployed" and those in the "inactive" categories reported that they also held a job, possibly because of the time difference between the *current* situation and *last week* (more likely in the case of those unemployed) and because the employment was not the main activity (more likely in the case of the inactive categories). Not surprisingly, a substantial proportion of those in full-time education also reported holding a job. Hence, for some groups, whether labour market survey questions ask for "main activity" or whether there was "any work" may be important for the resulting measures of employment. The main analysis uses only the "main activity" definition, but comparisons with the "any work" definition are presented in appendix C.

[30] The exception to this is students in interviews subsequent to wave A, for whom both sets of starting dates are collected but information on the previous spell corresponds to the break date for the main activity, leaving a potential gap in the "any work" history. In wave A, students with some employment did not record a start date for their main activity (education). In addition, in the data, some individuals reporting a non-work main activity and also some employment were found to have starting dates for their main activity identical to the starting dates for their employment, whereas the routing of the questionnaire indicates that they should be missing. The main activity starting dates for these cases were therefore set to missing.

[31] The base sample used in this paper consists of all adults (aged 16 plus) interviewed at any wave during the first nine waves, excluding proxy interviews (identified using the variable wivfio) and new households from the former ECHP sample introduced at the seventh wave (wave G) (identified using the variable wmemorig). Twenty cases with missing interview dates in wave A were also dropped.

In constructing the basic data set, five types of labour market state were distinguished: employment, self-employment, unemployment, out of the labour force and full-time education. Consecutive spells of the same type were merged together, with the exception of employment spells, which were divided by a change of employer[32], but not by promotion or job change within the same employer[33]. To create consistency across all data sources, dates were usually measured by months and years without days[34]. Several other minor technical adjustments were also made[35].

3. Recall Bias in the Wave Overlaps

3.1 Hypotheses on Recall Bias

When individuals are asked to recall events from the past, they may omit an event altogether, they may misclassify an event, or they may make a time error in the duration of the event or when the event occurred. Such errors may be made unintentionally or may be part of a respondent's conscious reinterpretation of the past. Several factors have been shown to influence the degree of recall error: the length of recall period; the complexity of the reporting task; and the saliency, pleasantness and social desirability of the events being recalled[36]. In terms of reporting past labour market activity, these factors suggest several hypotheses about recall error.

---

[32] For the "main activity" history, information from the wave data and from the main activity lifetime history collected in the second wave was used. Employer-to-employer moves from the any work lifetime history collected in the third wave were used to find the breaks between employers within employment spells for the second wave history data. Some 93 percent of these moves recorded in the third wave history could be matched into an employment spell in the second wave history. A high degree of consistency between the two lifetime histories was also found in (Halpin (1997) who compares the implicit monthly employment statuses between the two labour market histories (section 4.4.1).) In addition, the starting date for the current spell was set to missing for those with a non-work main activity who also reported some work. For the "any work" history, information from the any work history collected in the third wave was used alone with the wave data. In addition, the starting date for the current spell was set to missing for full-time students who also reported some work.

[33] In particular, spells of "different job/same employer" were merged with subsequent employment spells. Some 96 percent of such spells were followed by an employment spell and could be merged. One reason for ignoring the promotion or job change information was that the wording of the question asking employees for the date of last promotion/job change was altered after wave A with some apparent effect on responses. In wave A, the question was: "What was the date you started working in your present position, by that I mean the beginning of the current spell of the job you are doing now for your present employer ?" In subsequent waves, the following was inserted: "If you have been promoted or changed grades, please give me the date of that change."

[34] In some cases, seasons were reported in place of months. For the overlap data used in section 3, seasons were replaced as April (Spring) and July (Summer) and, in the years prior to interview year, as January (winter) and October (Autumn). Missing months in the years prior to the interview year were replaces as July. No adjustment was made for missing or season dates that could fall within the overlap period (autumn and winter of the interview year) as this could be crucial for the matching process between waves. For the history data used in section 4, the seasons were replaced as January (winter), April (Spring), July (Summer), October (Autumn) and missing months as July for years prior to 1990.

[35] First, any spells with starting dates after the interview date were dropped. Second, the ordering of the "main activity" options (variables ajbstat and ajhstat) in wave A was adjusted to match that in the subsequent waves. Third, the spell numbering in the main activity history in the second wave (variable bleshno) was adjusted to run consecutively. Fourth, the variable wnemst was used in preference to wjbstat where there was a conflict between the two in waves subsequent to A as the starting dates apply to this variable rather then wjbstat. Such a conflict arose in 1.2 percent of cases. Finally, as the employment history in the third wave collects information only on employers prior to the current employer, some current employment spells in wave A only have the start date for last promotion and not for time with current employer. Unless coinciding with the end of the last employment spell in the employment history or with the start of an employment spell in the main activity history from the second wave, the starting date for these cases were set to missing for the purposes of calculating employer tenure.

[36] See Sudman & Bradburn (1973), Akerlof & Yellen (1985), Dex (1995), Mathiowetz & Morganstein (1991).

First, fewer spells will be reported as the length of the recall period increases. Quite simply, events such as, jobs, periods of inactivity, or transitions between jobs or states, are less likely to be reported correctly as the length of time over which the individual must remember them increases. When constructing a labour market history, the omission of spells will implicitly subsume their time into other spells, while forgotten transitions will merge spells together. Although the distribution of aggregate time across labour market states may not be affected, the average duration of spells will increase and the frequency of transitions decline as the recall period lengthens.

Second, shorter spells are more likely to fail to be recalled than longer spells. Spells of shorter duration are typically less important than those lasting a longer time and are more easily forgotten. In addition, shorter spells mean more spells to recall, increasing the complexity of the reporting task and raising the potential for error.

Third, spells of unemployment are more likely to fail to be recalled than employment spells or spells of inactivity, even controlling for spell length. Salience of an event or spell depends upon whether anything memorable was happening during the period. Periods of unemployment may be less memorable, therefore, if they were simply time spent waiting for something else (a job) to begin. In addition, the unpleasantness or social undesirability of time spent looking for work may lead the respondent either to genuinely wipe such occurrences from memory or to consciously fail to reveal them. Indeed, the spell may be reclassified rather than forgotten altogether, particularly if it is a long spell. Respondents may also reinterpret the past to present an image more in line with what they felt to have occurred after the event. Consequently, spells of job search that did not result in employment but in exit from the labour force may be merged with the subsequent spell of inactivity. For all of these reasons, time in unemployment is less likely to be recalled correctly than periods of employment or inactivity.

Finally, the saliency, pleasantness and social desirability of different labour market spells may depend upon the type of respondent and recall bias may vary by individual characteristics. The salience of employment and unemployment may be greater for men than women because of the financial importance of being the prime household earner. On the other hand, time spent out of the labour force is likely to be more memorable for women than men because it is often connected with a positive and socially desirable role such as raising children. It has also been suggested that women are more likely to view themselves as out of the labour force rather than unemployed because they tend to be part-time and/or secondary workers. Consequently, women may be more likely to fail to recall spells of employment and unemployment than men or have a greater tendency to redefine unemployment spells as inactive, while men are more likely to fail to recall spells of inactivity. Similarly, the salience of employment and unemployment may be less for young and older age groups, who are also more likely to be part-time and/or secondary workers and whose

time in inactivity is likely to have a positive connection in the form of education for the young and retirement for the old. Hence, young and older age groups may be more likely to fail to recall spells of employment and unemployment than prime age groups, while prime age groups may have poorer memories for spells of inactivity.[37]

3.2 The Overlaps

Most interviews in the BHPS were conducted in the late autumn, but information on labour market spells is recorded from the 1st September of the previous year. Hence, there is a period of overlap between the waves when information for the same period is collected twice, initially with a recall period of a few months and then at the subsequent interview with a recall period of a few months plus one year. Comparing the initial report with the second report permits a measure of the impact of recall length on reported labour market behaviour[38].

In theory, the overlap period could begin with the starting date of the first spell recorded in the second report, but this would generate a bias in the nature of the changes between the two reports. In particular, it would tend to include cases where several short spells in the first report overlap with one long spell in the second, but would tend to exclude the symmetric case where one long spell in the first report overlaps with several short spells in the second[39]. Therefore, in order to capture a pure slice of labour market history without reference to spell transitions, the overlap period begins on the 1st September prior to the initial interview and ends on the day of that interview. The length of overlap ranges from 1 month to 9 months, with a median value of 2 months[40]. Spells falling within the overlap period were selected from each pair of consecutive waves[41]. Overlaps with incomplete spell types or incomplete or inconsistent dates were dropped from the sample[42].

---

[37] Levine (1993) finds that workers with relatively weak labour force attachment (including those whose main activity is "keeping house" or "in school") were less likely to report unemployment retrospectively. Evidence from the US presented in Mathiowetz & Duncan (1988) suggests that recall error may differ across demographic groups because of differences in salience and task measures rather than an inherent ability to recall unemployment correctly.

[38] The presumption is that the initial report, being closest in time to the event, is the most accurate account. This contrasts with the approach taken in Poterba & Summers (1986), where the second report is taken as the most accurate measure of status. This is because Poterba & Summers use "reconciled" data collected only a week after the initial report, designed specifically to obtain the "true" status and hence appropriate for measuring spurious transitions due to classification error in the initial report (although the accuracy of this reinterview data has been questioned (Biemer & Forsman (1992)). The second report in the BHPS data, on the other hand, uses information collected in essentially the same way as the initial report, the only difference being the additional year of recall. Hence, both reports may be subject to the same random misclassification, making the data unsuitable to compare the incidence of spurious transitions directly, but it is almost ideal for measuring the systematic impact of recall.

[39] For example, consider two hypothetical cases where the cut-off for questioning is month/year 9/90 in the first report and 9/91 in the second. In case (a), there are three first report spells beginning in 1/91, 3/91 and 5/91 and one second report spell beginning in month 1/91. In this case, the overlap begins in 1/91 and all four spells are included in the overlap. In case (b), there is one first report spell beginning in 1/91 and, although the individual actually remembers three spells beginning in 1/91, 3/91 and 5/91 at the time of second interview, only the last spell beginning on 9/91 is recorded in the second report. The overlap therefore begins in 5/91 and there is only one spell in both reports in the overlap.

[40] There were 61,987 overlaps with the following distribution of lengths: 39.0% (1 month), 41.3% (2 months), 14.7% (3 months), 2.6% (4 months), 1.1% (5 months), 0.6% (6 months), 0.5% (7 months), 0.2% (8 months) and 0.02% (9 months).

[41] Defining the overlap period and selecting the spells within the period was unique in using the day part of dates in order to ensure complete symmetry in selecting spells from both the initial and second report. In particular, it was important to ensure that

Each overlap was classified by the difference in the number of spells between reports and whether the sequence of spell types matched, ignoring, for the moment, whether the spell lengths matched. A summary of this classification is presented in table 2. Over 92 percent of all overlaps consist of one matched spell in both reports, while just under 1 percent have two or more spells in both reports which match in type. On the other hand, just over 4 percent have one spell in each report, but of different types. Overlaps are marginally more likely to have fewer spells in the second report (1.59 percent) than more spells (0.71 percent). Hence, the overwhelming majority of overlaps do not exhibit any recall error, which is not surprising given that the difference in recall is only one year. Moreover, the shortness of the overlap period makes it unlikely that either report will include more than one spell, substantially reducing the likelihood that there will be any change in the number of spells. In cases where there is more than one spell in at least one of the reports (3.52 percent of overlaps), the likelihood of the two reports not matching by type of spell is 74 percent, showing that recall errors are much likely when there are more transitions between labour market states. This is not surprising as there is more information to recall correctly, but these cases are also the more interesting ones for analysing labour market dynamics.

Recall errors are more likely among young respondents (about 14 percent of overlaps) than prime age (around 7 percent of overlaps) and older respondents (around 4 percent of overlaps), but there is no marked distinction between men and women. Figures not shown in the table indicate a propensity for recall errors to rise as the length of overlap increases, from around 5 percent for overlaps of one month to around 31 percent for overlaps of nine months[43]. This is not surprising as longer overlaps increase the opportunity for spells to disappear or appear. For this reason, overlap length is used as a control in the regressions for spell matching below. Interestingly, there was no pattern in the matching process across successive waves, suggesting that panel attrition or repeated interviewing do not affect recall reliability.

3.3 Matching Spells

In order to assess the effect of recall on particular types of spells and for specific kinds of respondents, a process was developed for pairing each spell in the first report with the best match that could be found in the second report. The criteria used for this pairing included whether spells overlap (they have at least one month in common), whether they match in type (employment, self-employment, unemployment, out of the

---

spells in the second report beginning in the interview month began before the interview day in order to be selected into the overlap.

[42] 1446 (2.3 percent) of the overlaps were dropped for this reason.

[43] The percentage of overlaps not matched (defined as the same number and type of spells) is 5.4 for an overlap length of one month, 6.9 for two months, 8.5 for three months, 11.4 for four months, 12.2 for five months, 13.0 for six months, 12.4 for seven months, 9.3 percent for eight months and 30.8 percent for nine months.

labour force or full-time education) and whether they match in position (1st spell, 2nd spell, etc in the overlap). The pairing procedure can be summarised[44]:

1. Spells were paired if they overlapped, were of the same type and were in the same position.
2. Of the remaining unpaired spells, spells were paired if they overlapped and were of the same type, but not in the same position[45].
3. Of the remaining unpaired spells, spells were paired if they overlapped and were in the same position, but were not of the same type.
4. Of the remaining unpaired spells, spells were paired if they overlapped, but were of different types and positions[46].

Hence, priority was given to finding overlapping spells of the same state. Allowing the end dates of spells to be truncated at the first interview date, spells in the initial report could be defined according to their pairing with spells from the latter report:

- perfect match: spell type and start and end dates identical
- shifted match: spell type and length identical, but dates different
- lengthened match: spell type identical, but spell longer in second report
- shortened match: spell type identical, but spell shorter in second report
- redefined: spell types different
- disappeared: spell in first report is not paired

The "disappeared" spells in the initial report must overlap with spells in the second report that have been paired with other spells from the initial report. They can therefore be analysed by the type of spell(s) that they have disappeared into, including different spells of the same type. The "redefined" spells can also be analysed by the type of spell they are redefined into. Spells that are not paired in the second report are defined as having "appeared". Examples of this matching procedure are provided in appendix A.

A summary of the proportions of spells in the initial reports that are matched, redefined and disappear in the second report is presented in table 3a. Spells of employment are most likely to be recalled a year later, with around 97 percent of spells in the initial report being matched with spells in the latter report for prime age men and women. On the other hand, spells of unemployment are least likely to be

[44] In the actual program, there was an initial step that paired spells where both reports contained a single spell (captured in steps (1) and (3)) and the listed steps were only applied to overlaps with more than one spell in at least one report.
[45] This does not necessarily give a unique pairing, so the order of pairing involved looking for a pair for the first spell in the initial report among all the spells in the second report, then for the second spell in the initial report, and so on.
[46] Again, the pairing is not necessarily unique and the ordering was the same as step (2).

recalled correctly a year later[47], with as few as 52 percent matched in the second report for young and prime age women[48]. The ranking of other spell types depends upon the type of respondent. Young men and women are particularly good at recalling full-time education spells (91 percent are correctly matched) relative to older groups, but their recall of employment, self-employment and out of the labour force spells (for young men) is weaker than for older groups. Old men and women rarely fail to recall spells out of the labour force (99 percent of spells are correctly matched), but are less likely to recall spells of unemployment than younger groups (only 66 percent of spells for old men and 40 percent of spells for old women are correctly matched). Finally, although men and women have similar matching rates for many types of spells, women are much less likely to recall unemployment spells correctly than men. For example, prime age women recall only 52 percent of unemployment spells a year later compared to 74 percent for prime age men. It is noticeable that these patterns suggest that spells are more likely to be recalled correctly when they are the more common behaviour for a particular group.

Tables 3b through 3f provide detailed breakdowns of the matching for each spell type. The last row of each table presents the average change in spell length for matched spells. The first of these tables, for employment, shows that even when the matching rate is very high, reporting the same spell starting date in both reports is much less common. Even allowing dates to be measured by the month (rather than the day), less than half of the employment spells are recalled with the same starting month a year later. This "perfect matching" is slightly lower for old men and women than for younger groups, possibly reflecting longer spells and the need to recall the starting date over a longer period. It is interesting to note that the tendencies to lengthen or shorten the reported spell length with an additional year of recall are roughly equal. However, the mean change in spell length in the final row of the table shows that, on average, the spell length has been shortened by about a month for young and prime age individuals, but lengthened by almost 5 months for old men. In terms of changing the reported type of activity, employment spells are most likely to be redefined as or disappear into spells of full-time education and unemployment for young men and women and into spells out of the labour force for old individuals. Women of all ages are more likely than men to redefine employment spells into time out of the labour force rather than into unemployment[49]. It is interesting to note that this redefinition tends to be towards labour market activity that is more common for the group. There is also some tendency for employment spells to be subsumed into

[47] This is consistent with Poterba & Summers (1986) who find that misclassification errors are much more common for those in unemployment than those in employment or out of the labour force (table III).

[48] The lowest matching rate is actually for full-time education spells for old women, but the sample size for this group and for education spells for old men is too small to draw reliable conclusions.

[49] Poterba & Summers (1995) find a corresponding result in the CPS data in that employed women are more likely than employed men to misclassify themselves as out of the labour force (table 2).

other spells of employment over the longer recall period, particularly for young individuals[50].

The matching patterns for spells of self-employment (shown in table 3c) exhibit similar rates for spells being lengthened and shortened as for employment spells. In addition, there is no significant change in the mean spell length with the additional year of recall, suggesting that although there may be considerable recall error in reporting the starting date, this error is random and may average out over larger samples. However, the likelihood of a perfect match is lower for spells of self-employment than employment because a substantial proportion of the spells are redefined as employment or disappear into spells of employment. Indeed, almost 22 percent of self-employment spells reported by young men in the initial interview have been redefined as employment or disappeared into employment at the following interview. For women of all ages, the proportion is around 10 percent, although the fraction is much lower at approximately 4 percent for prime age and old men. Again, the likelihood of such redefinition appears to be related to the proportions of a group in a particular labour market state: individuals in groups with a higher fraction in self-employment are less likely to recall a different state a year later. There is also a tendency for time in self-employment to be redefined as unemployed time for young men and women.

The likelihood that initial reports of spells of unemployment will be perfectly matched in a subsequent report a year later is low for all groups, ranging from 16 percent for old women to 36 percent for old men (table 3d). The proportions being lengthened and shortened are, however, fairly evenly balanced and there is no significant change in the reported spell length[51], as was the case with self-employment spells. For young men and women and prime age men, the low matching rate for unemployment spells is mostly accounted for by spells being redefined as or disappearing into employment spells (22 percent, 22 percent and 12 percent of spells respectively). Substantial proportions are also redefined as and disappear into time out of the labour force for young women and prime age men. On the other hand, the low matching rate is *mostly* accounted for by redefinition and disappearance into time out of the labour force for prime age women and old men and women (31 percent, 29

[50] Recall that these employment spells are defined by change in employer rather than simply promotion or position change for a given employer.

[51] Contrasting results are found for the US in Bowers & Horvath (1984) and Poterba & Summers (1984). Bowers & Horvath, using monthly interview data from May-August 1976 from the CPS, find that, among the continuously unemployed, there is a tendency for the reported duration of unemployment to increase by more than the time lapsed between surveys. The proportions losing weeks, giving consistent responses and gaining weeks are 38, 24 and 37 percent for May-June, 38, 27 and 35 percent for June-July and 45, 25 and 30 for July-August, while the average errors are 1.72, 0.87 and 0.16 for the three pairs of months (table 1). However, they also find that there is a considerable amount of overstatement of change among shorter spells, while there was increasing understatement among longer spells. Poterba & Summers compare reports of unemployment duration in similarly matched files from the CPS for May and June 1976. They find that 32 percent gave consistent reports of the duration, while 32 percent reported a shorter duration in June (allowing for the fact that the duration should have increased by 4-5 weeks) and 37 percent reported a longer duration, although the magnitude of the overstatement tended to be greater than the understatement (table 2). Demographic group was not important in the likelihood of error, except for teenage women who systematically underreport the duration increment (table 3).

percent and 47 percent respectively). Indeed, these represent substantial switches from unemployment into inactivity for these groups[52].

For spells out of the labour force, the perfect matching rates generally lie between the low levels for unemployment and the high levels for employment, with the proportion of spells being lengthened and shortened once again being reasonably balanced (table 3e). The only significant change in spell length is for prime age women, for whom the average spell length increases by almost 2 months between reports. Spells not matched tend to be redefined as or disappear fairly equally into spells of employment or unemployment.

For spells of full-time education (table 3f), the tendency for the length of spell to alter with the additional year of recall is similar to those for other types of spells, but greater proportions are shortened than lengthened. Indeed, for young men, the average spell length declines by almost 10 months between reports while the fall is over 13 months for young women. Large inconsistencies in the reporting of the these spell lengths may not seem so surprising when it is noted that the start dates for most spells are referring to the time the individual started school as a young child, although why there should be such a marked change in one direction so many years after the original date is not clear. Spells of full-time education that are not matched tend to be redefined as or disappear into employment.

These differences in the impact of recall by type of spell and respondent group may be partly explained by spell length if shorter spells are less likely to be recalled correctly than longer ones. In order to assess the impact of spell length on recall error, multinomial logit models for the probabilities of match, redefinition and disappearance for initial spells were estimated. These models included regressors for spell length interacted with each spell type as well as dummy variables for spell type (with employment as the omitted variable)[53] and variables for overlap length. Overlap length was also included in the regressions to control for the fact that shorter overlaps are more likely to generate matches or redefinitions than disappearances. Separate models were estimated for men and women, but not for the age groups due to an absence of variation in some outcomes once controls for spell length were included.

The results for the two regressions are shown in table 4. For men and women, the likelihood of redefinition and disappearance declines with spell length (shown by relative risk ratios significantly less than one) for all spell types except for the redefinition of unemployment and education spells. This provides strong evidence that shorter spells are more likely to be forgotten than longer spells as the length of

[52] Poterba & Summers find corresponding results in the CPS data that unemployed women are more likely than unemployed men to misclassify themselves as out of the labour force (table III in (1986) and table 2 in (1995)).

[53] Without the dummy variables for spell type, the slopes with respect to spell length would reflect the change from the *average* matching rate for all spell types at zero length rather than the desired measure from zero length for each spell type. The dummy variables therefore reflect differences across spell type at zero length allowing individual slope parameters, which is not the same as the differences controlling for spell length. The differences controlling for spell length are captured in the model shown in table 6.

recall increases, possibly due to their lower salience. In addition, the rate of decline in the probability is greater for disappearance than for redefinition, although this may be due to the shortness of the overlap period and the fact that only spells in the initial report that do not cover the entire overlap period can possibly be categorised as having disappeared. Indeed, it is not surprising that while the overlap length does not significantly affect the probability of redefinition, the likelihood of disappearance increases significantly with the overlap length. For women, the likelihood that a spell of unemployment will be redefined actually rises with spell length. As shown above (table 3d), the vast majority of these spells are redefined as time out of the labour force, suggesting that while short unpleasant or socially undesirable unemployment spells may be conveniently forgotten, longer spells may be merged into a subsequent spells of inactivity or find justification in activities outside of the labour market[54].

In order to obtain an idea of the magnitude of the changes in the likelihood of redefinition and disappearance as spell length increases, tables 5a and 5b present the predicted probabilities at 3 months, 12 months and 60 months for men and women respectively. For men and women, the likelihood of redefinition declines at a steady rate for employment, self-employment and spells out of the labour force. On the other hand, the probability that a spell of unemployment will be redefined rises dramatically for women, from 29 percent at 3 months length to 50 percent at 60 months. The likelihood of disappearance drops markedly between 3 months and 12 months length for all spell types except for time out of the labour force for women and education spells for both men and women.

How much of the difference in recall error across spell type and respondent group can be explained by variation in spell length ? To address this question, a single logit model was estimated for the probability of a spell being matched which included control variables for the length of spell and overlap length (linear and squared terms). The model also included 28 dummy variables for spell type, gender and age interactions, excluding old men and women in full-time education[55]. The percentages of spells that the model predicts to be matched at the median spell length (52 months) and median overlap length (2 months) for each spell type, gender and age category are presented in table 6. The significance of differences between these predictions is measured as the significance of the difference between the corresponding coefficients in the model.

---

[54] Mathiowetz & Duncan (1988) and Levine (1993) find that shorter spells of unemployment are less likely to be recalled than longer ones in the US. Mathiowetz and Duncan compare survey reports with company records of unemployment in the prior two years and find that 25 percent of spells of less than one week were reported in the interview compared to 34, 39, 43, 56, 51 percent for spell lengths of 2, 3-4, 5-12, 13-20, 21-28 and 29+ weeks (table 2). Levine, using CPS data, finds that shorter in-progress spells of unemployment are less likely to be reported retrospectively the following year than longer ones. For example, in 1987, 60 percent of spells with an in-progress length of 4 weeks or less fail to be reported retrospectively compared to 42 percent for those with length 4 to 16 weeks and 40 percent for those over 16 weeks (table 3).

[55] The spell type full-time education was omitted for old men and women due to the small number of observations in each category (6 and 4 respectively).

The effects of spell type and respondent group on the spell matching rate controlling for spell and overlap length are broadly similar to the raw averages shown in table 3a. The most notable difference is that, although spells of unemployment are still the least likely type of spell to be matched, the size of the difference is smaller, showing that part, but not all, of the poor recall for unemployment can be explained by the shortness of spells. The final row of table 6 shows that many of the differences between spell types are significant, with the main exceptions being between spells out of the labour force and education spells and between self-employment and education for young women and prime men. Within spell types, most differences in matching rates across age groups are significant, the main exceptions being between the young and old for employment spells and between young and prime age individuals for unemployment and education. Significant differences in the matching rate between men and women are rarer: there are significant gender gaps for spells of unemployment for all age groups, but only for prime age men and women for employment and self-employment and only for young men and women for spells out of the labour force. Nevertheless, the earlier conclusions regarding differences in recall by spell type and respondent category are shown to hold to a significant degree and are not explained by differences in spell length.

3.4 Aggregate Labour Market Activity

In order to assess whether recall error may lead to systematic bias or simply introduce a random measurement error, the effects of certain types of spells being redefined and forgotten need to be balanced against the likelihood that spells of the same type will be added through redefinition and the appearance of new spells. Two aggregate measures of labour market behaviour are therefore compared between initial and second reports.

The first measure is the distribution of time use between the five types of spell. The impact of recall varies considerably across respondent gender and age (table 7). For all respondent groups, there is a significant fall in the reported time spent in unemployment with the additional year of recall. The magnitude of the fall in terms of the ratio between the second and initial report is greatest for old men and women, with women experiencing larger declines than men in all age categories. For example, the reported proportion of time in unemployment is only 91.7 percent of that in the initial report for prime age women compared to 93.2 percent for prime age men[56]. For

[56] Much of the previous literature on recall bias has been analysed in terms of unemployment rates or proportions of time in a particular state and most of the findings are consistent with those in table 7. Using evidence from Britain, Elias (1997) compares unemployment rates reported retrospectively in the BHPS lifetime histories with contemporaneous reports in the Labour Force Survey and finds that underreporting in the BHPS (up to three years earlier relative to the LFS) is most severe for prime-age women (table 1). Dex & McCulloch (1998) compare reports of unemployment in the BHPS wave B lifetime history with those from a similar retrospective survey (Family and Working Lives Survey) and find greater discrepancies for women than men in the unemployment rate between the two surveys. For the US, Morgenstern & Barrett (1974), Horvath (1982), Levine (1993) and Akerlof & Yellen (1985) use CPS data to compare the current report of unemployment with the retrospective previous calendar

young men and women, there is also a significant decline in the reported time spent in full-time education (and in self-employment for young men), while time in employment and time out of the labour force are the net beneficiaries from the additional year of recall. For prime and old men, the decline in reported time in unemployment is accounted for by a rise in time out of the labour force. Old men also report significantly less time in self-employment after an additional year of recall.[57]

The second aggregate measure is the number of spells of labour market activity. The net changes in the number of spells (shown in table 8) suggest that the recall effects are not random. The total number of spells reported declines significantly for all gender and age groups (except old men), ranging from a fall of 3.5 percent for young men to 0.2 percent for old women. As was the case with the distribution of time use, the largest changes are for young men and women. The number of unemployment spells decline considerably (by 18 percent for men and 24 percent for women), while there is also a decline in the number of education spells (by 3 percent) and a sizable drop in self-employment spells for young men (13 percent). However, there are also rises in the number of spells out of the labour force by 24 percent for young men and 5 percent for young women. For prime and old groups, the largest changes are also in the number of unemployment spells, ranging from a decline of 9 percent for prime age women to a fall of 30 percent for old women. Prime age women also report significantly fewer spells in employment and self-employment after an additional year of recall, while old men report fewer self-employment spells after the longer period. There are net rises in the number of spells only for spells out of the labour force for prime and old men.

Hence, most of the net changes in the reported distribution of time use are mirrored by net changes in the reported number of spells: fewer unemployment spells for all groups, fewer education spells for the young and fewer self-employment spells for young and old men on the one hand, and greater numbers of spells out of the labour force for the young individuals and for prime and old men on the other hand. In contrast, the greater proportion of time in employment in the second report for

---

year report in the March Work Experience Supplement and find substantial understatement in the retrospective reporting of time in unemployment. Levine finds that differences in retrospective and contemporaneous in the CPS for 1988 are highest for teenage white men (9.3 and 13.9 percent) and women (6.6 and 12.3 percent), but are greater for white prime age women (3.2 and 4.2 percent) than for similar men (4.0 and 4.3 percent) (table 2, 1988). Ratios of retrospective to contemporaneous reports from the CPS for 1960-81 are presented in Akerlof & Yellen as 0.780, 0.614, 1.151, 0.889, 1.161 and 1.186 for young men and women, prime men and women and old men and women respectively (table III). Morgenstern & Barrett use similar data for 1964-71 and find that the retrospective measure is more likely to understate the contemporaneous measure for women (by 24 percent for white females) than for men (3 percent for white males). They also find that understatement is greater for men and women aged under 24 than for older age groups. In contrast, Duncan & Hill (1985) and Mathiowetz & Duncan (1988) compare administrative data with survey data for a large manufacturing company in the US and find little bias in the reporting of unemployment in the previous year or in current employer tenure.

[57] Previous work on the US has suggested that a longer recall period will lengthen the amount of time in employment. Powers et al (1978) analyse data from a survey of males in nonmetropolitan areas in Iowa and find that two thirds recalled 10 years later that they had worked more weeks in 1964 than they reported at the time. Ryscavage & Coder (1989) compare data from the CPS WES and SIPP and find that the shorter recall period in the SIPP generates a smaller proportion of the workforce employed in full-time year-round employment.

young men and women is not reflected in a significant rise in the number of spells of employment, suggesting that there is a net lengthening of employment spells.

3.5 Summary

The evidence suggests that most individuals give consistent reports when asked to recall their labour market behaviour a year later. However, this conclusion is based on a relatively short window of comparison that provides a comparatively small opportunity for recall error. Indeed, the accuracy of second reports is much lower for individuals with more than one spell to report. In addition, the reported starting dates for all types of spells exhibit much lower levels of consistency than just recalling the spell. Although there is a fairly even balance between the proportion of spells that are lengthened and the proportion that are shortened, the average length of matched employment spells declines between the two reports for young and prime age individuals, while the mean spell length rises for old men in employment and for prime age women out of the labour force

In spite of this high degree of consistency, most of the hypotheses about recall bias are confirmed. Significantly fewer spells are reported with the additional year of recall than in the initial accounts. Shorter spells are significantly more likely to be redefined or disappear than longer spells, with the exception that the probability of redefinition rises with spell length for unemployment for women. Even controlling for differences in spell length, spells of unemployment are less likely to be recalled correctly than other types of spells, with the consequence that the reported time in unemployment and number of unemployment spells decline significantly as the recall period lengthens. In addition, spells of self-employment are particularly likely to be redefined as time in employment, with less time in and fewer spells of self-employment over the longer recall period. On the other hand, time out of the labour force is greater and the number of spells in such inactivity is higher with the additional year of recall.

The evidence generally supports the hypothesised effects of recall for different gender and age groups. In particular, women are less likely to correctly recall spells of unemployment than men, with a greater tendency to redefine time as out of the labour force. Old men and women also have a higher likelihood than younger groups to fail to recall unemployment spells and to redefine time as being spent out of the labour force. Consequently, the decline in unemployment time and spells as recall lengthens is greatest for these groups. The consistency in reporting employment and self-employment spells is particularly poor for the young, but there is also a relatively high tendency among the young to redefine unemployment as employment. The net effect of the additional year of recall for this group is a decline in time and the number of spells in education and a rise in the proportion of time in employment.

4. Labour Market Histories

Having documented the evidence that recalling labour market behaviour over even as short a period as one year may lead to marked biases in recorded responses, the next question to ask is whether and how much these biases matter in analysing dynamic patterns of employment ? In addition, do they generate substantially worse or better degrees of accuracy than the potential bias of spurious transitions from repeated interviews ? In order to attempt to answer these questions, labour market histories were constructed for the BHPS data using different methods that allow the potential biases to influence the final measures to differing degrees.

4.1 Alternative Construction Methods

In contrast to the wave overlap analysis, the labour history data was constructed using three (rather than five) labour market states: employment, unemployment and out of the labour force. Spells of self-employment were treated as an employment spell (effectively for the employer "self"), while spells of full-time education were not analysed and individuals dropped if a period of analysis required including time in full-time education. The main reason for the use of only three states was to allow comparisons with previous work on labour market dynamics that has tended to focus on just the three categories. Labour market histories were constructed for each individual interviewed at each wave, going back to the date the individual last left full-time education whenever possible. All four methods potentially draw on information from the current and all prior interviews, but deal with the inconsistencies generated by the wave overlaps in different ways:

- Method A: Traditional Panel: Ignores any information collected for the previous interview date or time prior to that date.

- Method B: Latest Interview Rules: Only uses information from prior interviews for the time not covered by the current interview.

- Method C: Reconciled: Uses amendments to resolve as many inconsistencies in the wave overlaps as possible and to impute missing information.

- Method D: Selected No Problems: Selects observations with no inconsistencies or missing information.

Method A gives complete priority to information from the initial, closest, interview rather than from the second interview for time falling in the wave overlap. It

is the method closest to a "traditional panel" in the sense that measurement errors in the current state at the time of interview may generate spurious transitions[58]. This method minimizes the impact of recall bias, although it should be noted that some of the data is still collected on a retrospective basis and the potential for recall error has not been completely removed. Method B effectively uses information from the second interview (or accounts from even later interviews) for the time falling in the wave overlap and the output corresponds most closely to that which would be collected in a retrospective history. However, rather than collecting data for the entire potential labour market history at a single interview, information at a particular interview is only collected as far back as the start of the spell covering the 1st September of the previous year and prior information is then completed from the closest interview. Hence, this method is an amalgam of a pure retrospective history and a panel survey. Method C is similar to the traditional panel method in giving priority to information from the closest interview, but also uses a series of amendments to remove inconsistencies between reports from different interviews, including those leading to spurious changes immediately after interviews[59]. By dropping from the sample those cases where the inconsistencies could not be resolved, this method has the lowest likelihood of spurious transition bias as well as a minimum degree of recall bias. In this sense, it may generate the most "accurate" picture of the labour market history and is the baseline from which the impact of spurious transition bias (as captured in method A) and recall bias (as captured in method B) is measured[60].

Method D considers a slightly different approach to the problem. In deriving data samples for analysis is it not uncommon to simply select observations with complete and consistent data and it is useful to consider the impact that such an

---

[58] The construction of the BHPS Combined Work-Life History Data deposited in the ESRC Data Archive was based on similar principles to method A (Halpin (1997), section 3). In particular, inconsistencies were resolved by allowing an earlier account, being nearer the time described, to overrule any later report. In addition, the histories were initially constructed in a monthly calendar format and then converted to spell-type data. The potential for seam effects was explicitly recognized, with two suggested solutions (section 4.3). The first is to estimate Cox proportional hazard models that include a variable for a transition at a seam (termed the "transition effect"), which shows that seams are strongly associated with transitions. In addition, it is shown that failure to include this control has a substantive influence on the other parameter estimates in the model, indicating that the seam problem will bias model estimates if not suitably addressed. The second suggested solution for coping with the seam effect is to randomise the transition to lie at some point between the two interview dates, but this does not appear to have been incorporated in the data. This second suggested solution is similar to the approach used in method C in this paper to deal with the spurious transitions. The deposited data is analysed in Oskrochi & Crouchley (2000), who find that employment transitions tend to bunch in the interview months (as would be expected) and that there are inconsistencies in other variables across datasets.

[59] The amendments are listed in appendix B. They were derived by examining case-by-case inconsistencies and missing information and were designed to generate consistency for a high number of cases without making imputations that dramatically changed the data. Hence, the listed amendment rules are not comprehensive in that some theoretical possibilities did not arise in the data used here. It should be noted that these amendments had to be consistent with all other information in the history for the observation to remain in the sample.

[60] Oskrochi & Crouchley (2000) describe an alternative method to construct the lifetime employment histories. They use amendments similar to those described for method C to ensure internal consistency, but only within each data source (that is, from each wave and from the lifetime histories collected in waves B and C). When combining the sources, the primary (determining) source was generally the data set that was collected as soon as possible after the event of interest, including giving priority to the wave B history over the C history, as in method A. However, the wave B lifetime history was allowed to overrule anything collected in wave A and current status at any interview time was overruled by any error-free spell from wJOBHIST from the subsequent wave within 24 months. This latter criteria was designed to reduce seam effects and operates in a similar manner to the amendments for inconsistencies in method C. Hence, it might be expected that the dataset generated by Oskrochi & Crouchley would have similar properties in terms of biases as the "most accurate" reconciled (method C) dataset used here.

approach would have for the analysis of labour market dynamics. Method D therefore also minimizes the spurious transition and recall biases as in method C, but consists of a very different sample. Indeed, the method is a subset of the output from each of the first three methods because the cases selected with no inconsistencies or missing information should have identical histories no matter what the method of construction.

In drawing comparisons between each of these methods, it is useful to distinguish between differences arising from "sample selection" and those from "construction". Selection differences will arise because some methods will generate a valid and complete history for a particular observation while other methods may not. In particular, method C will draw additional observations into the sample that have incomplete or inconsistent data under the other methods but are reconciled in this approach. In addition, there may be "construction" differences for the same observation when different construction methods are applied. Constraining the samples to include only those observations for which there is a valid and complete history from the methods being compared reveals these pure construction effects. However, only the first three methods can be compared for these construction differences, as any observations from method D will generate identical histories under any construction method.

The notes to the tables below show that the sample sizes differ considerably across methods, generating potentially large impacts for the selection effect. Consistent with prior expectations, the "selected no problem" method generates the smallest sample (it is a subset of all other methods), while the "traditional panel" produces the second smallest sample (because it draws only on one source of information for each point in time). The "latest interview rules" creates the second largest samples (because it can draw on secondary sources of information where it is missing), while the "reconciled" method has the largest sample size (because it draws on all information and reconciles). Indeed, the "reconciled" sample is often considerably larger than those for the other methods, showing how successful it can be in resolving inconsistencies that arise outside of the wave overlaps.

4.2 Division of Time

The impact of the different methods on the amount of time spent in each of the three states is presented in table 9. The top panel for each state presents the mean amount of time spent in that state for each of the gender and age groups over the 8 to 24 months prior to interview. Differences across methods capture both sample selection effects and pure construction impacts. The lower panel for each state presents the differences in the proportion of time across paired observations and captures the impact of the construction method without any sample selection effects. The period of 8-24 months prior to interview was chosen for the comparison in order

to maximises the potential impact of the differences in the construction methods and the effects of the biases. In particular, the 8 months immediately prior to interview would be identical for all methods as they draw on the same source of information, while the period prior to this takes in the wave overlap where differences will occur. Considering a longer period might begin to blunt the differences in recall while also reducing the number of valid observations[61].

The table highlights the fact that women spend a much smaller proportion of their time in employment and unemployment than men and experience more time out of the labour force. For example, prime age women spend around three-quarters of their time in employment compared to about 90 percent for prime age men. Somewhat surprisingly, the gender differences are almost as marked for young men and women, although the differences amongst the older group are less unexpected given the difference in retirement age and cohort effects on the propensity of women to work. Age is also an important determinant of labour market status. The proportion of time spent in unemployment declines with age category, while men and women in the oldest category spend the vast majority of their time out of the labour force.

Looking across the different construction methods, the most marked contrasts in the raw (unpaired) means are between the approach that selects those observations with completely consistent reports across interviews (method D) and the reconciliation approach (method C). In particular, the former generates much lower unemployment rates than the latter. For example, young men are reported to spend 13.6 percent of their time in unemployment under method C in comparison to 9.7 percent method D. For the young and prime age individuals, the selected no problems method (method D) generates the highest amount of time in employment (typically, 3 to 4 percentage points higher than any other method) and the lowest proportion of time out of the labour force. The reconciliation method (method C) generates a sample with the lowest proportion of time spent in employment for young and prime age individuals and some of the highest proportions of time out of the labour force. On the other hand, for those in the old age category, the reconciled method generates the highest proportions in employment, while the sample of those without any inconsistencies generates the lowest proportions of time spent in employment, although the difference is much more marked for old women than old men.

Many of these contrasts in the raw means reflect the selection effects of the different methods. In particular, method D tends to select individuals with few transitions between states for whom it is easier to give consistent responses, while method C, by reconciling the problematic cases, tends to add reports with greater numbers of movements back to the sample. Hence, the latter method is more likely to include those with some (typically temporary) time in unemployment. In addition, if

[61] However, looking back over longer periods, up to 10 years prior to interview, generated very similar patterns and conclusions to those presented in table 9.

employment is a more transient state for the old age group than for the younger age groups, the reconciliation method is likely to generate a greater proportion of workers in its sample of old respondents. On the other hand, the selected no problems approach is likely to have a disproportionate share of more stable individuals, including those working in the young and prime age groups and those out of the labour force in the old category.

Examining the differences in paired observations removes the sample selection effects and reduces many of the differences between methods A to C[62]. The latest interview rules method (method B) stands out as generating significantly different results from the other two results, while the traditional panel (method A) and reconciliation (method C) approaches differ little in the measured proportions. For the time in employment, the proportions generated using the latest interview rules approach are significantly greater for young and prime age men and significantly smaller for old men than for the other methods. However, using this method generates significantly less time in unemployment for all categories of respondents than other methods, while generating greater time out of the labour force for some groups. Given that method B allows the greatest role for recall error, these differences suggest that recall bias significantly increases the amount of reported time in employment for young and prime men; raises the reported time out of the labour force for young women, prime men and old individuals; and reduces the time in unemployment for virtually all groups. On the other hand, the consistency between the traditional panel and reconciliation methods provides support for the hypothesis that misclassification errors in current state at the time of interview are essentially random and do not have a significant impact on the aggregate proportion of time spent in each state.

4.3 Transitions Between Labour Market States

This subsection turns the analysis towards the measurement of *dynamic* labour market statistics with a consideration of monthly transition rates. Tables 10a through 10c present the percentage of those in a particular state in an initial month who transit to a different state in the following month or, in the case of the first panel in table 10a, who move to a different employer. The period covered is, again, 8 to 24 months prior to the interview in order to maximise the potential impact of both biases. In contrasting the transition rates between methods, only the combined impact of selection and construction is analysed as the level of observation is the month and generating "matched" observations by months would be complicated (as they would not always be of the same state) and the conclusions difficult to interpret[63].

[62] As mentioned above, the contrasts between the selected no problems method (method D) and the other methods cannot be analysed because this method selects only those observations with identical labour histories across all methods.
[63] The test of significance used for differences between the methods is for differences in the mean for two unmatched samples.

Looking across all three tables, it is clear that the likelihood of a transition varies considerably by the type of change. The most likely move is from unemployment into employment, while a change in the opposite direction or from employer to employer are also more common types of movements. The probability of a change of state also depends upon the gender and age group of the respondent. For those in employment, women are more likely than men to transit to a different employer or to move out of the labour force in the subsequent month, the latter difference being particularly marked for young and prime age women. On the other hand, men in employment are more likely to move into unemployment in the following month than their female counterparts. For those in unemployment, women are more likely than men to transit into employment or out of the labour force in the subsequent month, particularly, again, young and prime aged women. Women out of the labour force are less likely than their male counterparts to move into unemployment. However, while young women are less likely to transit from out of the labour force into employment than young men, prime age women are more likely to do so than prime age men. Young men and women are generally more likely to make a transition of any type than prime age respondents, while those in the old category tend to be least likely to make any type of move. The only exception to this pattern is that old men and women are more likely to move from employment to out of the labour force than both the younger groups, presumably reflecting retirement from paid work.

There is a very distinct pattern in these transition rates across the different methods of constructing the labour market history. With the exception of unemployment to employment transitions[64], the traditional panel approach (method A) and the reconciliation approach (method C) generate significantly higher transition rates than the latest interview rules approach (method B) and the selected, no problems approach (method D). These differences are often substantial. For example, an average 2.0 and 1.9 percent of young men in employment in an initial month transit to a different employer in the subsequent month according to the traditional panel and reconciled methods respectively, compared to 1.6 percent for the selected no problems method and 1.5 percent for the latest interview rules method (table 10a). For prime age women in unemployment, 2.8 percent transit out of the labour force in the subsequent month according to the traditional panel and reconciled methods, compared to 0.6 and 0.5 percent according to the latest interview rules and selected no problem methods respectively (table 10b). These differences across the methods are not surprising. The selected no problem sample (method D) is most likely to drop those with frequent moves who are most likely to report inconsistencies in their behaviour.

[64] Quite why the transition from unemployment to employment does not follow the same pattern as all other transitions is not immediately obvious (top panel in table 10b). It may relate to particularly small sample size for this category, especially as the one other case where the pattern across methods does not hold (for transitions from out of the labour force to employment for young men) also has an unusually small sample size.

Although it is not possible to separate selection effects from pure construction impacts, fewer transitions under method B are consistent with recall bias reducing the number of reported transitions as some spells are forgotten or subsumed into others. What is more surprising is the size of this difference: the combined effects of selection and construction algorithm have a substantial impact on the dynamics measured in the data. However, there is a lack of a consistent substantial difference between the traditional panel and reconciliation methods and, in many cases, transition rates are actually significantly higher for the latter. This may be due to sample selection: although the reconciliation method may remove spurious transitions in the period immediately after interview, it may also draw in observations which are particularly dynamic in their labour market behaviour by resolving inconsistencies at other times Hence, there is no conclusive evidence that spurious transition bias has an impact on measured transition rates[65].

4.4 Survival Models for Spells of Labour Market Activity

Survival models were estimated for each state, using all spells that appear within a consistent segment of a respondent's labour market history[66]. Spells from the entire labour market history were considered to illustrate the impacts of the biases over the longer period. Once again, only the combined impacts of variation in selection and construction could be tested as matching spells across methods was not always feasible. Indeed, it could be expected that an important source of differences in spell lengths and survival rates would arise from variation in the spells reported. For each state and gender and age group, a Weibull survival model was estimated with a series of three dummy variables for each method (method A being the omitted category). The model accounted for the fact that some spells were censored in their length at the time of interview. The median survival time for each method was then predicted from the estimated coefficients[67]. Three tests of the significance of differences in the distribution of spell lengths across methods were considered: testing differences in the coefficients in the Weibull model and log-rank and Wilcoxon/Breslow tests for the equality of survivor functions across methods. The predicted median spell lengths from the Weibull models and the test results are shown in table 11.

The median spell length with a given employer is slightly longer for young women than young men, but prime and older women generally have shorter employer tenures than their male counterparts. Consistent with the gender differences in

[65] This contrasts with the results shown in table V in Poterba & Summers (1986). Although based on annual rather than monthly transition rates, the table shows consistently lower transition rates for all types of changes when the data is adjusted for spurious transitions caused by current state classification error. However, there is no sample selection issue in their data when making the comparisons.

[66] Working backwards from the interview, all spells are included in the sample until there was an inconsistency.

[67] Using the median time as a point of comparison, as well as removing the impact of outlying observations, captures differences in the degree of skewness in the survival distribution as well as the mean duration.

transition rates out of unemployment, women of all ages have shorter median spells lengths in unemployment than their male counterparts. For time out of the labour force, young and prime age women have much longer median spell lengths, although there is little difference in the length for men and women in the old age category. Across all types of spell, the median spell length increases substantially with age group. For employment and unemployment spells, this is likely to reflect the fact that older respondents have longer labour market histories and therefore the potential for longer spells. For time out of the labour force, the longer spells for old respondents reflect the permanency of the state of retirement.

The pattern in median spell length across difference methods of construction is very distinct: the last interview rules approach generates significantly longer median spells lengths across all three types of spell and most gender and age groups, while the traditional panel approach creates the shortest median spell length in most cases[68]. For example, the median employment spell length for prime age men is 42.3 months according to the last interview rules method and 31.8 months according to the traditional panel approach. The median unemployment spell length for prime age women is 6.4 months according to the former method and 4.7 months according to the latter. For young women, the last interview rules approach generates a median spell length out of the labour force of 22.9 months compared to 15.6 months for the traditional panel approach. The other two methods (reconciliation and selected no problems) generate median lengths that lie in between these two extremes, but are not always significantly different from the traditional panel approach.

Although selection and construction effects cannot be separately identified, the evidence is consistent with the hypothesis that recall bias generates significantly longer spell durations for all three types of spells. In addition, although the pattern is not as strong or consistent, the fact that the traditional panel method generates shorter, and often significantly shorter, spells than the reconciliation method is consistent with the idea that spurious transition bias may lead to shorter measured spell lengths.

4.5 Wage Returns to Experience and Tenure

Finally, the employment history data is combined with wage data to illustrate how the biases may be related to other key variables. The level of observation is the individual at each interview and contrasts can therefore be drawn between the impact of selection and the effect of the construction algorithm across methods.

Average years of employment experience (including both paid employment and self-employment) for all respondents, regardless of current labour market status, are presented in table 12. There is little difference in experience between young men and

[68] This is consistent with the evidence presented for expected unemployment durations in table VII in Poterba & Summers (1986) which shows that failure to adjust for spurious transitions caused by classification errors in current status generates substantially shorter spells of unemployment for all gender and age categories of respondents.

women, but prime age and old women have spent considerably fewer years in employment than their male counterparts[69]. For example, while prime age men have spent an average 21 years in work, prime age women have spent an average 16 years.

Examining the average experience across the construction methods shows that there is significant variation between the methods, but that there is no consistent pattern of differences. For example, the reconciliation method (method C) generates the highest levels of experience for young and old respondents, but the lowest level for prime age individuals, while the selected no problems approach generates the lowest levels for young men and women and the highest levels for prime age individuals[70]. In addition, in some cases, the combined selection and construction effects shown in the top panel in table 12 generate a different ranking to the construction impact alone shown in the bottom panel. For example, for the unmatched data for old women, average experience is 28.02 years under the traditional panel approach and 28.13 years under the reconciliation method, but, for the matched observations, average experience is still 28.02 years under the former method but significantly lower at 27.96 years under the latter method. Given that selection into the sample requires a consistent account of experience since leaving full-time education, the selection effect may be playing an important, but complicated, role in the differences. Yet even for the paired observations, there is no obvious pattern across the methods.

Average employer tenure, defined as the number of years with the same employer, for those currently in paid employment (but not self-employment) is presented in table 13. Women have shorter average tenures than men[71]. For example, prime age women average between 5 to 6 years of tenure compared to 7 to 8 years for men. As would be expected, the average tenure also rises across age groups. It is not surprising, given the analysis of spell lengths above, that the latest interview rules approach (method B) to constructing the labour market histories generates significantly higher levels of tenure than any other method, while the traditional panel approach (method A) creates the shortest tenures. The discrepancies between methods are slightly greater in the top panel of table 13 than in the paired differences in the bottom panel, showing that both selection effects and the construction algorithms are contributing towards the differences across approaches. This evidence suggests that even when the sample is held constant, recall bias generates significantly and

[69] Manning & Robinson (1998) show that men and women in the first five waves of the BHPS who are currently in employment have very similar levels of experience (table 4a). The contrast with the figures here might be expected from conditioning upon those currently in employment.

[70] Given the differences in the division of time for the 8-24 months preceding the interview detailed in section 5.2 above, it might have been expected that the selected no problems method (method 7) would have generated the higher levels of experience for young and prime age individuals, while the traditional panel approach (method 5) would have led to the greatest years of experience for old men and women.

[71] Manning & Robinson (1998) also report that current job tenure is higher for men than women in data from the first five waves of the BHPS (table 4a).

substantially longer measures of employer tenure, while spurious transition bias significantly reduces the average length.

The final tables (14a and 14b) show how these differences in measured experience and tenure interact with current wage levels[72] for men and women currently in paid employment[73]. The top row presents four basic wage regressions, one for each method, while subsequent rows present the coefficients on interaction terms added to the basic model for one comparison method in term (generating a total of six comparison regressions, each with one set of interactions terms). The significance of these interaction terms show whether the coefficients from the basic models in the top row differ significantly across method. The log wage level is estimated as a quadratic function of both experience and employer tenure. Although the quadratic is not the most ideal fit for this type of model, the main aim is to compare results across different methods for constructing the experience and tenure variables and the quadratic is convenient for this purpose[74].

For men, the linear element of the estimated return to employer tenure lies between 1.4 percent and 1.9 percent each year, while the linear component of the return to experience is much higher, ranging from 5.3 percent to 5.5 percent each year (table 14a). In both cases, significant negative coefficients for the quadratic terms show that these returns declines as years of tenure or experience accumulate. Looking across the base models in the top row of table 14a, it can be seen that the highest returns to tenure are generated by the traditional panel and reconciliation approaches (methods A and C) and the lowest by the latest interview rules approach (method B). This is consistent with the traditional panel having the lowest measured average levels of tenure and the latest interview rules the highest levels of tenure. The differences in the estimated returns to experience across methods are much smaller, but the latest interview rules approach (method B) generates the lowest return. Turning to the interaction terms in the other rows of table 14a, it can be seen that the main significant difference in returns is the higher return to tenure for the reconciliation method over the latest interview rules approach (method C over method B). This suggests that recall bias may lead to a significant underestimate of the returns to tenure for men. There are also significant differences in the overall model estimates (including the constant term) between the reconciliation method on the one hand and the traditional panel and selected no problems approaches on the other.

For women, the linear element of the estimated return to employer tenure is higher than that for men, lying between 2.0 percent and 2.6 percent each year, while the linear component of the return to experience is much lower than that for men,

[72] The wage measure is the gross hourly wage including usual overtime, indexed to autumn 1999.
[73] The sample sizes were too small to divide the analysis by age group.
[74] For similar reasons, no other variables are included in the regression.

ranging from 2.2 percent to 2.7 percent each year (table 14b)[75]. Again, these returns decline as tenure or experience accumulate. The ranking in the return to tenure across methods is similar to that for men with the traditional panel and reconciliation approaches (methods A and C) generating the highest returns and the latest interview rules (method B) the lowest return. The difference between the traditional panel and latest interview rules approaches are also significant, indicating that data subject to recall biases may generate significantly different results from that subject to spurious transition bias. The differences across methods in the return to experience have the same ranking for women as for men, but are much greater in magnitude for women. Indeed, the latest interview rules approach (method B) generates a significantly lower return than the reconciliation and selected no problems approaches (methods C and D), suggesting that recall bias may lead to underestimates of the return to experience for women. As was the case for men, there is also significant a difference in the overall model estimates between the reconciliation method and the selected no problems approach.

4.6 Summary

The analysis above confirms previous findings that there are substantial differences in labour market behaviour between men and women and across age groups. In terms of the division of time, women are more likely than men to be found out of the labour force, while average time in unemployment declines with age category. Young men and women are those most likely to move from one labour market state to another, while men and women in the oldest age group are the most stable in their behaviour. Women in employment are more likely than men to move to another employer or to exit the labour force, while employed men have a greater probability of transiting into unemployment. Unemployed women are also more likely to move into employment or out of the labour force than unemployed men. The probability that women out of the labour force will move into unemployment is lower than that for similar men. Young women are less likely than young men to move into employment from out of the labour force, while prime age women are more likely than similar men to make the transition. Relatedly, women have shorter median spell lengths than men for employment and unemployment, but young and prime women have longer median spell lengths for time out of the labour force than their male counterparts. Not surprisingly given the differences in the division of time use,

[75] Manning (1998a) finds similar patterns in the returns to experience for men and women using data from the General Household Survey (table 1). In addition, Manning & Robinson (1998) use data from the first five waves of the BHPS to estimate the returns to experience and tenure allowing for a wide variety of controls. Using a quartic specification for experience, their results suggest similar profiles for men and women for an initial 10 years, but then a more rapid decline in the returns for women than men. The estimated returns to tenure (using a quadratic) are smaller than those reported here, which is not surprising given the additional control variables in the regression, but the estimated return for women is considerably higher than that for men (table 5). Using NLS data from the US in similar regressions to those used here, Light & Ureta (1995) find that women have lower returns to experience than men, but higher returns to tenure (actual experience regressions in table 3).

women have considerably lower levels of accumulated employment experience than men. Finally, women have higher wage returns to tenure than men, but lower returns to experience.

The evidence presented above has also shown that the method used to construct data on labour market histories can be influential on measured behaviour. In particular, the method that allows the greatest impact for recall bias generates significantly higher proportions of time in employment for young and prime age men; significantly greater proportions of time out of the labour force for young women, prime age men and old individuals; and significantly less time in unemployment for all age and gender groups. Recall bias is also shown to produce data with lower transition rates between states and significantly longer median spell lengths across all types of spells, while average employer tenure is also higher. Moreover, the evidence suggests that recall bias in the experience and tenure data is likely to lead to significant underestimates of the wage returns to tenure for men and of the returns to experience for women.

On the other hand, there is no evidence that spurious transition bias affects on the division of time, as would be expected, or, more surprisingly, transition rates between labour market states. The bias does, however, generate generally shorter median spell lengths and lower levels of average employer tenure. It should be noted that the effect of classification error in generating spurious transitions has been considered only in an indirect manner, comparing measured labour dynamics using only a single original report of current status with a "most accurate" picture based on a reconciled account from two reports, both of which may contain classification error and the latter of which is reported one year after the event. Hence, the lack of any evidence should not be interpreted as evidence of an absence of bias, particularly with regards to the impact on transition rates.

Selecting a data set on the basis of observations with consistent data has also been shown to generate a distorted picture of the labour market. Such an approach generates lower unemployment rates for all gender and age groups, while raising the measured time in employment for young and prime age individuals and increasing the time out of the labour force for old men and women. From a dynamic perspective, using the selection approach generates lower transition rates between states, although there is no evidence of a substantial impact on median spell lengths.

5. Conclusions

Correctly measuring individual dynamics in labour market behaviour has become increasingly important as research and policy attention has become more focused on the relationships between current employment opportunities and past

experience. Surveys collecting information on labour market histories use repeated interviews and retrospective reporting, laying the resulting data open to potential biases from spurious transitions due to random measurement errors and from systematic recall error.

This paper has considerably extended the existing knowledge of the nature of the bias in retrospectively recalling labour market spells. It has shown that most individuals are consistent in their reporting as the recall period lengthens, but individuals with the most transient behaviour are more likely to give inconsistent accounts. There is a considerable degree of inconsistency in the reporting of spell start dates, but the likelihood of spells being lengthened and shortened with additional recall length is fairly evenly balanced. Somewhat surprisingly, there is a tendency for employment spells to be shortened for young and prime age individuals, but old men tend to lengthen employment spells while prime age women tend to lengthen spells out of the labour force. There is no evidence to support previous findings that the reported length of unemployment spells increases with recall.

Many of the hypotheses concerning recall bias are confirmed by the data. Fewer spells are reported as the recall period lengthens. Previous work has suggested that shorter spells of unemployment are less likely to be recalled correctly than longer ones, but the evidence presented here shows that this is true for all types of spells including employment, self-employment, out of the labour force and education spells. The one exception to this pattern is that women are more likely to redefine unemployment spells as time out of the labour force as spell length increases. This paper confirms earlier results that aggregate time in unemployment declines with recall, but it also shows that that number of unemployment spells declines significantly and that unemployment spells are less likely to be recalled correctly than other types of spells even allowing for differences in spell length. In addition, recall is shown to have similar impacts on self-employment, much of which is due to self-employment spells being redefined as paid employment. But recall has opposite impacts on time out of the labour force: aggregate time and the number of spells rise. The hypotheses concerning differences in recall bias across gender and age groups are generally supported by the data. Previous results that the under-reporting of unemployment is most severe for women and for younger groups is confirmed. In addition, this paper has shown that unemployment time tends to be redefined as time out of the labour force for women and old individuals, while the young tend to redefine the unemployed time as employment.

The analysis of recall bias has also demonstrated some new results on the measurement of labour market dynamics. In particular, it has shown that the bias significantly reduces transition rates between all types of states and increases median spell lengths for employment, unemployment and time out of the labour force. A longer recall period generates higher average levels of employer tenure for those in

paid employment and can affect estimates of wage returns to tenure and experience. The evidence presented here suggests that recall bias may lead to underestimates of the return to tenure for men and of the return to experience for women.

The analysis of the impact of spurious transition bias is not as conclusive as that for recall, mainly because the effects could not be tested so directly. As would be expected, the bias does not significantly alter aggregate measures of the division of labour market time between states. The evidence on transition rates between states does not support previous findings that the bias will lead to an overstatement of the degree of dynamics. But the examination of median spell lengths confirms earlier work that spurious transitions will significantly shorten spell lengths for all types of spells, while the bias is also shown to reduce the average reported length of employer tenure.

Whether these biases generate "large" differences in the resulting estimates depends upon the analyst's perspective and the precise motivation for measuring labour market dynamics. In particular, it is difficult to answer conclusively which of the two biases is greatest and hence whether there should be any preference in collecting labour market history data between repeated interviews and retrospective reporting. To what degree other survey data will be prone to similar biases will depend upon the specifics of the survey, particularly the frequency of the interviews and length of recall period. Given the careful design of the BHPS, it would be surprising if the biases in other data sources were substantially smaller than those evidenced here.

The value of the information presented in this paper is that it allows analysts to judge whether and to what degree their conclusions are likely to be compromised by the reporting biases. For example, using lifetime retrospective data to conclude that labour market transition rates have risen over time could be questioned on the grounds that recall bias alone could generate such a picture without any real change in behaviour. More usefully, the knowledge of how the biases operate permits some results to stand unchallenged by concerns over the accuracy of the data reporting. For example, using the same type of dataset to conclude that the likelihood of self-employment declines over the lifetime could not be doubted as an artefact of recall bias: indeed, recall bias would only serve to moderate the measured change. Hence, while the analysis of the biases cannot draw a *general* judgement that one type of data collection method is superior to another, it can provide, for *particular* research questions, some guidance on the type of data source that is least likely to generate biased results through recall or measurement error.

References

Akerlof, G. A. and Yellen, J. L. (1985). 'Unemployment Through the Filter of Memory', *Quarterly Journal of Economics*, August, pp. 747-773

Altonji, J. and Williams, N. (1998). 'The Effects of Labour Market Experience, Job Seniority and Mobility on Wage Growth', *Research in Labor Economics*, 17, pp. 233-276

Arulampalam, W. (2001). 'Is Unemployment Really Scarring ? Effects of Unemployment Experience on Wages', *The Economic Journal*, vol. 111(November), pp. F585-606

Arulampalam, W., Booth, A. and Taylor, M. P. (2000). 'Unemployment Persistence', *Oxford Economic Papers*, vol. 52, 24-50

Arulampalam, W., Gregg, P. and Gregory, M. (2001). 'Unemployment Scarring', *The Economic Journal*, vol. 111(November), pp. F577-584

Biemer, P. P. and Forsman, G. (1992). 'On the Quality of Reinterview Data with Application to the Current Population Survey', *Journal of American Statistical Association*, vol. 87(420), pp. 915-923

Booth, A. L. and Francesconi, M. (2000). 'Job Mobility in 1990s Britain: Does Gender Matter ?', *Research in Labour Economics*, 19, pp. 173-189

Booth, A. L., Francesconi, M. and Garcia-Serrano, C. (1999). 'Job Tenure and Job Mobility in Britain', *Industrial and Labour Relations Review*, vol. 53(1), pp 43-70

Booth, A. L. and Frank, J. (1996). 'Seniority, Earnings and Unions', *Economica*, vol. 63, pp. 673-86

Booth, A. L., Jenkins, S. P. and Garcia-Serrano, C. (1999). 'New Men and New Women ? A Comparison of Paid Work Propensities from a Panel Data Perspective', *Oxford Bulletin of Economics and Statistics*, vol. 61(2), pp.167-197

Bowers, N. and Horvath, F. W. (1984). 'Keeping Time: An Analysis of Errors in the Measurement of Unemployment Duration', *Journal of Business and Economic Statistics*, vol. 2(2), pp. 140-149

Bowlus, A. J. (1997). 'A Search Interpretation of Male-Female Wage Differentials' *Journal of Labor Economics*, vol.15(4), pp. 625-657

Bronars, S. and Famulari, M. (1997). 'Wage, Tenure and Wage Growth Variation Within and Across Establishments', *Journal of Labor Economics*, vol. 15(2), pp. 285-317

Brown, J. N. and Light, A. (1992). 'Interpreting Panel Data on Job Tenure', *Journal of Labor Economics*, vol.10(3), pp. 219-257

Clark, K. B. and Summers, L. H. (1979). 'Labor Market Dynamics and Unemployment: A Reconsideration', *Brookings Papers on Economic Activity*, vol. 1, pp. 13-72

Dex, S. (1995). 'The Reliability of Recall Data: A Literature Review', *Bulletin de Methodologie Sociologique*, no. 49, pp 58-89

Dex, S. and McCulloch, A. (1998). 'The Reliability of Retrospective Unemployment History Data', *Work, Employment and Society*, vol. 12(3), pp. 497-509

Duncan, G. J. and Hill, D. H. (1985). 'An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data', *Journal of Labor Economics*, vol.3(4), pp. 508-532

Dustmann, C. & Meghir, C. (2001). 'Wages, Experience and Seniority', Institute for Fiscal Studies Working Paper 01/01

Eisenhower, D., Mathiowetz, N. A. and Morganstein, D. (1991). 'Recall Error: Sources and Bias Reduction Techniques', in (P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz and S. Sudman, eds) *Measurement Errors in Surveys*, New York: Wiley

Elias, P. (1997). 'Who Forgot They Were Unemployed ?', Working Paper Number 97-19, ESRC Research Centre on Micro-Social Change, University of Essex

Flinn, C. J. and Heckman, J. J. (1983). 'Are Unemployment and Out of the Labor Force Behaviorally Distinct Labor Force States ?', *Journal of Labor Economics* vol. 1(1), pp. 28-42

Gonul F. (1992). 'New Evidence on Whether Unemployment and Out of the Labour Force are Distinct States', *Journal of Human Resources*, vol. 27(2), pp.329-361

Gregg, P. (2001). 'The Impact of Youth Unemployment on Adult Unemployment in the NCDS', *The Economic Journal*, vol. 111(November), pp. F626-653

Gregory, M. and Jukes, R. (2001). `Unemployment and Subsequent Earnings: Estimating Scarring Among British Men', *The Economic Journal*, vol. 111(November), pp. F607-625

Halpin, B. (1997). 'Unified BHPS Work-Life Histories: Combining Multiple Sources into a User-friendly Format', *Technical Papers of the ESRC Research Centre on Micro-Social Change*, Technical Paper 13, Colchester: University of Essex

Harkness, S. (1996). 'The Gender Earnings Gap: Some Evidence for the UK', *Fiscal Studies*, vol 17(2), pp. 1-36

Holt, D., McDonald, J. W. and Skinner, C. J. (1991). 'The Effect of Measurement Error on Event History Analysis', in (P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz and S. Sudman, eds) *Measurement Errors in Surveys*, New York: Wiley

Horvath, F. W. (1982). 'Forgotten Unemployment: Recall Bias in Retrospective Data', *Monthly Labor Review*, March, pp. 40-43

Hujer, R and Schneider, H. (1989). 'The Analysis of Labor Market Mobility Using Panel Data', *European Economic Review*, vol. 33, pp. 530-536

Jones, S. and Riddell, C. (1995). 'The Measurement of Labor Force Dynamics with Longitudinal Data: The Labour Market Activity Survey Filter', *Journal of Labor Economics*, vol. 13(2), pp. 351-385

Kalton, G., Kasprzyk, D. and McMillen, D. B. (1989). 'Nonsampling Errors in Panel Surveys', in (D. Kasprzyk, G. Duncan, G. Kalton & M. P. Singh, eds) *Panel Surveys*, New York: Wiley

Kletzer, L. G. (1998). 'Job Displacement', *Journal of Economic Perspectives*, vol. 12, pp. 115-36

Levine, P. B. (1993). 'CPS Contemporaneous and Retrospective Unemployment Compared', *Monthly Labor Review*, August, pp. 33-39

Light, A. and Ureta, M. (1995). 'Early-Career Work Experience and Gender Wage Differentials', *Journal of Labor Economics*, vol. 13(1), pp. 121-153

Machin, S. and Manning, A. (1999). 'The Causes and Consequences of Long-Term Unemployment in Europe', in (O. Ashenfelter and D. Card, eds) *The Handbook of Labor Economics*, volume 3, Amsterdam: Elsevier Science

Manning, A. (1998a). 'Movin' On Up: Interpreting the Earnings-Experience Profile', London School of Economics, Centre for Economic Performance Discussion Paper no. 380, January

Manning, A. (1998b). 'Mighty Good Thing: The Returns to Tenure', London School of Economics, Centre for Economic Performance Discussion Paper no. 383, March

Manning, A. and Robinson, H. (1998). 'Something in the Way She Moves: A Fresh Look at an Old Gap', London School of Economics, Centre for Economic Performance Discussion Paper no. 389, May

Mathiowetz, N. A. and Duncan, G. J. (1988). 'Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment', *Journal of Business and Economic Statistics*, vol. 6(2), pp. 221-229

Morgenstern, R. D. and Barrett, N. S. (1974). 'The Retrospective Bias in Unemployment Reporting by Sex, Race and Age', *Journal of the American Statistical Association*, vol. 69(346), pp. 355-357

Myck, M. and Paull, G. (2001). 'The Role of Employment Experience in Explaining the Gender Wage Gap', Institute for Fiscal Studies Working Paper 01/18

Oskrochi, G. and Crouchley, R. (2000). 'Using SPSS and Gauss to Unify the BHPS Work History Data', Centre for Applied Statistics Working Paper no. 2000/03, University of Lancaster

Paull, G. (1997). 'Dynamic Labour Market Behaviour in the British Household Panel Survey: The Effects of Recall Bias and Panel Attrition', The Labour Market Consequences of Technical and Structural Change Discussion Paper Series no. 10, Centre for Economic Performance

Peracchi, F. and Welch, F. (1995). 'How Representative are Matched Cross-Sections ? Evidence from the Current Population Survey', *Journal of Econometrics*, vol. 68, pp. 153-179

Poterba, J. M. and Summers, L. H. (1984). 'Response Variation in the CPS: Caveats for the Unemployment Analyst', *Monthly Labor Review*, March, pp. 37-43.

Poterba, J. M. and Summers, L. H. (1986). 'Reporting Errors and Labor Market Dynamics', *Econometrica*, vol. 54(6), pp.1319-1338.

Poterba, J. M. and Summers, L. H. (1995). 'Unemployment Benefits and Labor Market Transitions: A Multinomial Logit Model with Errors in Classification', *Review of Economics & Statistic*, vol. 77(2), pp. 207-216

Powers, E. A., Goudy, W. J. and Keith, P. M. (1978). 'Congruence Between Panel and Recall Data in Longitudinal Research', *Public Opinion Quarterly*, vol. 42, pp. 380-389

Ryscavage, P. & Coder, J. (19890. 'Longitudinal vs. Retrospective Measures of Work Experience', Survey of Income and Program Participation Working Paper 8903, Washington D.C.: US Bureau of the Census

Stevens, A. H. (1997). 'Persistent Effects of Job Displacement: The Importance of Multiple Job Losses', *Journal of Labor Economics*, vol. 15, pp. 165-88

Sudman, S. and Bradburn, N. M. (1973). 'Effects of Time and Memory Factors on Response in Surveys', *Journal of the American Statistical Association*, vol. 68(344), pp. 805-815

Appendix A: Matching Spells Examples

The bold notation shows the start dates and spell type for each spell reported.

Example 1: Perfect Match Prime male, wave B, interview in Sep 92	
1st report	**Oct 88: emp**
2nd report	**Oct 88: emp**
Matching	perfect match

Example 2: Shifted Match Young male, wave E, interview in Dec 95				
1st report	**July 95: emp**	**Sep 95: unemp**	**Sep 95: emp**	**Oct 95: emp**
2nd report	**Aug 95: emp**			**Oct 95: emp**
Matching	shifted match	disapp. into emp	disapp. into emp	perfect match

Example 3: Lengthened Match Young female, wave E, interview in Oct 94			
1st report	**Sep 95: emp**	**Nov 95: emp**	**Dec 95: unemp**
2nd report	**Sep 95: emp**	**Nov 95: unemp**	
Matching	perfect match	disappeared into unemp	lengthened match

Example 4: Shortened Match Prime female, wave D, interview in Dec 95	
1st report	**July 73: out of labour force**
2nd report	**July 74: out of labour force**
Matching	shortened match

Example 5: Redefinition from Education to Employment Young male, wave A, interview in Nov 91		
1st report	**May 91: emp**	**Oct 91: educ**
2nd report	**May 91: emp**	**Nov 91: emp**
Matching	lengthened match	redefined as emp

Example 6: Redefinition from Unemployment to Employment Young male, wave G, interview in Oct 97				
1st report	**May 96: emp**		**Sep 97: unemp**	**Oct 97: emp**
2nd report	**March 96: emp**		**Sep 97: emp**	**Oct 97: emp**
Matching	lengthened match		redef. as emp	perfect match

Example 7: Disappearance Prime female, wave F, interview in Oct 96			
1st report	**Aug 96: out of labour force**	**Sep 96: emp**	**Oct 96: out of labour force**
2nd report	**Aug 96: out of labour force**		
Matching	lengthened match	disappeared into out of labour force	disappeared into another spell of out of labour force

Example 8: Disappearance Young female, wave E, interview in Dec 95					
1st report	**Aug 89: emp**	**Oct 95: out of lab. force**	**Oct 95: educ**	**Nov 95: unemp**	**Dec 95: emp**
2nd report	**Aug 89: emp**	**Oct 95: unemp**			**Dec 95: emp**
Matching	perfect match	disapp. into emp & unemp	disapp. into emp	length. match	perfect match

Example 9: Appearance Prime male, wave C, interview in March 94						
1st report	**Sep 92: unemp**		**Oct 93: emp**	**Nov 93: unemp**		
2nd report		**Sep 93: emp**	**Oct 93: unemp**	**Nov 93: emp**	**Nov 93: unemp**	**Dec 93: emp**
Matching for 1st report	redefined as emp		redef. as unemp	shortened match		
Matching for 2nd report		redef. as emp	redef. as unemp	added spell	short. match	added spell

Example 10: Appearance Old female, wave G, interview in Sep 97			
1st report		**July 97: unemp**	
2nd report	**Mar 97: self-emp**		**Sep 97: educ**
Matching for 1st report		redefined as self-emp	
Matching for 2nd report	redefined as self-emp		added spell

Appendix B: Amendments in the Reconciliation Method

Some amendments corrected for inconsistencies in spell dates (where spells overlapped or there were gaps between spells):

a) Inconsistent months originally imputed from a reported season were adjusted (within the season) to match the reported month from another interview.

Example: A spell is reported to begin in winter 1988 (imputed as January) at the first interview and then in December 1988 at a subsequent interview. The start date is set as December 1988.

b) End dates for spells with a gap before the following spell which are in the same or consecutive waves were extended to the start date of that following spell

Example: Spell A ends in October 1995 and the subsequent spell B starts in February 1996. The end date for spell A is set as February 1996.

c) The start date for the second spell in overlapping spells was set to the end date of the first spell. If the spells had the same start date or the first spell began after the second, the second spell is dropped.

Example 1: Spell A starts in October 1993 and ends in March 1995. Spell B starts in January 1995 and ends in December 1997. The start date for spell B is set to March 1995.

Example 2: Spell A starts in July 1991 and ends in August 1995. Spell B starts in April 1992 and ends in August 1995. Spell B is dropped.

Other amendments imputed missing dates:

a) Missing start dates for the first spell after leaving full-time education were replaced with the date left full-time education if the spell ended within 12 months of leaving full-time education.

Example: The date left full-time education is August 1986 and the start date for the first spell is missing while the end date is February 1987. The start date for the first spell is set as August 1986.

b) Missing "dividing" dates between two spells (that is, where the start date for the first spell is known but the end date unknown and where the end date for the second spell is known but the start date unknown) were replaced with the midpoint of the feasible period for the start date.

Example 1: Spell A begins in February 1992 and is followed by spell B that ends in June 1992, but the break date between the two spells is missing. The feasible period for the missing start date for spell B is March to May and the start date is set as April 1992.

Example 2: Spell A begins in February 1992 and is followed by spell B that ends in January 1993, but the break date between the two spells is missing. However, at the time of interview (October 1992), the individual is in spell B. Hence, the feasible period for the start date of spell B is March to September and the start date is set as June 1992.

c) Missing start dates were replaced as the previous spell's end date where the end dates of the two spells were within 12 months of each other and the spells were of different types.

Example: Spell A of unemployment ends in March 1997 and spell B of employment ends in October 1997, but has an unknown start date. The start date for spell B is set as March 1997.

d) Unemployment and out of labour force spells with missing end dates were merged with subsequent spells of the same type.

Example: Spell A is unemployment with a start date of January 1993 but unknown end date. Spell B is also unemployment with a start date of October 1993 and end date of December 1994. The two spells are combined as a single spell of unemployment starting in January 1993 and ending in December 1994.

e) Employment spells with missing end dates were merged with subsequent spells of employment if the subsequent spell started at the same time or before the initial spell.

Example: Spell A starts in October 1995 but the end date is unknown. Spell B of the same type starts in September 1995 and ends in July 1997. The spells are combined as a single spell starting in October 1995 and ending in July 1997.

f) Spells with missing start dates were merged with previous spells of the same type if the end dates were within 12 months of each other.

Example: Spell A of employment has an unknown start date and an end date of June 1994. Spell B is also employment with a start date of March 1989 and an end date of June 1994. The two spells are combined as a single spell of employment starting in March 1989 and ending in June 1994.

Other amendments imputed missing spell types:

a) Spells of unknown type were merged with overlapping spells of known type.

Example: Spell A of unknown type starts in March 1993 and ends in April 1994. Spell B of known type starts in May 1993 and ends in March 1994. The spells are combined as a single spell of type B starting in March 1993 and ending in March 1994.

b) Spells of unknown type with either start or end date missing were merged with spells of known type if either the start or end dates were within 3 months of each other.

Example: Spell A of unknown type and missing start date ends in June 1992. Spell B of known type starts in May 1991 and ends in July 1992. The spells are combined as a single spell of type B starting in May 1991 and ending in July 1992.

Appendix C: Comparisons with the "Any Work" Definition of Employment

The tables and main text of the paper have considered the impact of recall for labour market activity measured using the "main activity" definition, but it is useful to consider whether recall has a similar effect if the "any work" definition is used.

For the spell matching in the wave overlaps (section 3.3), using the alternative "any work" definition generates few differences in the spell matching. There is poorer matching of self-employment spells for women and fewer education spells are reported in both the initial report and second report. However, there are some notable differences in the impact of recall on the aggregate labour market statistics in the wave overlaps (section 3.4). In particular, using the "any work" definition reduces the degree of recall error in the distribution of time use for young men, removing any significant difference between the initial and second report for the proportion of time spent in self-employment and education. On the other hand, the alternative definition increases the degree of recall error for women, generating a larger decline in the proportion of time spent in employment for the longer recall period and a larger rise in the proportions of time spent out of the labour force and in education. The latter result is not surprising: if such spells out of the labour force and in education tend to contain more minor spells of employment for women, then using the "any work" definition which may detect such current minor spells but not in retrospect will increase the impact of the longer recall length. This may also partly explain why the alternative definition leads to a greater decline in the number of employment and self-employment spells reported with the longer recall period for young and prime age individuals. In addition, the generally lower reporting of education spells appears to remove the negative impact of the additional year of recall on the number of spells of education recorded.

The differences between the two employment definitions are more noticeable in the construction of the labour market histories (section 4). Not surprisingly, with the exception of two cases, the "any work" definition of employment leads to higher reported amounts of time in employment over the previous 8-24 months than the "main activity", with the greatest differences for women and old men and the smallest discrepancy for young and prime age men (section 4.2). Although the proportion of time in unemployment is not greatly affected by the choice of employment definition, less time is reported out of the labour force when the "any work" definition of labour market behaviour is used rather than the "main activity" definition, particularly for prime age women and old men and women. These groups are those most likely to view employment as a secondary activity (family cares being women's main activity and retirement being the main activity for old individuals) and are therefore to be expected to be those with the greatest contrast across employment definition. The contrast between the two employment definitions in the division of time is greatest for the traditional panel and selected no problems approaches when there are both sample selection and construction differences. However, once the sample selection effect is removed, using the traditional panel and reconciliation methods generate the largest significant differences between the two definitions, while the selected no problems method produces almost identical proportions across the two definitions. One explanation for the traditional panel and reconciliation approaches to lead to the greatest differences is that both methods give priority to information given at the time of interview when both the "main activity" and "any work" definitions of employment choices are collected, rather than allowing later information to sweep back over the interview, blurring the impact of clashes between the two employment definitions.

There were few significant differences between the employment definitions for the transition rates (section 4.3). Using the "any work" definition rather than the "main activity" definition of employment generated a generally higher employer-to-employer transition rate. This may be explained by a greater propensity to move employers among those who regard their employment as secondary to another activity. The differences are largest and significant for the methods more sensitive to the distinction in definition (the traditional panel and the reconciliation approaches) and for the groups with the largest sample sizes (prime age men and women). The selected no problems method also detected the difference to a slightly lesser degree.

Differences between the two employment definitions were quite marked for the survival models (section 4.4). For employment spells, the "any work" definition significantly reduced the median spell length for prime age and old men for most methods, which is consistent with the notion that it is drawing in "secondary" shorter employment spells, particularly for old men. There are also significant differences in median spell length across the definition for prime and old women, but the median length fell under some methods and rose under others, suggesting that such "secondary" employment for women may not be of such a markedly shorter nature than primary jobs. For spells of unemployment, the employment definition had little consistent impact on the measured median spell length. For time out of the labour force, use of the "any work" definition generated generally significantly longer median spell lengths for prime age men and women, while the increases were both significant and substantial for old men and women. For example, using the "main activity" definition for the latest interview rules approach for old men generates an estimated median spell length of 219 months, compared to 1895 months for the "any work" definition. Underlying this extremely large change is a selection effect: although the "any work" definition has a higher mean and median spell length in the raw data than the corresponding "main activity" data, it has many fewer spells. In particular, the "main activity" sample has many more shorter, uncensored spells, which may not appear in the "any work" sample because they also correspond with some secondary employment. Hence, it appears that the definition of employment is especially important not just for estimating survival models for employment spells themselves, but also for modelling spells of other types that may coincide with some "secondary" employment.

Using the "any work" definition of employment in place of the "main activity" definition might be expected to generate higher levels of experience by adding time in employment even when not regarded as the main activity (section 4.4). However, this is only the case for the reconciliation approach (method C). For other methods, even for the matched observations, the total employment experience is typically significantly lower when the "any work" definition is used rather than the "main activity" which is puzzling. Using the "any work" definition of employment rather than the "main activity" definition generates substantially lower levels of tenure. This may be due in part to a selection effect whereby the "any work" definition adds secondary jobs of shorter tenure to the sample. However, the definition also generates significantly shorter average tenures between paired observations suggesting that there is an element in the construction algorithm that creates the difference. For men and women, using the "any work" definition of employment rather than the "main activity" generates significantly higher returns to experience for all methods (except the reconciliation method for men) and significantly higher returns to tenure for women with the reconciliation method.

Table 1: Current "Main Activity" and Work

Percentage of Individuals in Group	Young (under 25 years)		Prime (25 – 54 years)		Old (over 54 years)	
	Men	Women	Men	Women	Men	Women
In paid employment	49.8	45.2	71.5	66.1	19.7	15.1
Maternity leave	0.0	0.1	0.0	0.0	0.0	0.0
Maternity leave plus job	0.0	0.5	0.0	0.9	0.0	0.0
Government training	0.5	0.3	0.2	0.1	0.0	0.0
Government training plus job	1.2	0.8	0.0	0.0	0.0	0.0
Total in Paid Employment	**51.5**	**47.0**	**71.8**	**67.1**	**19.7**	**15.1**
Total Self-Employed	**2.9**	**1.0**	**15.3**	**5.2**	**8.7**	**1.9**
Unemployed	11.0	6.1	6.4	2.5	2.8	0.7
Unemployed plus job	0.6	0.4	0.3	0.2	0.1	0.0
Total Unemployed	**11.6**	**6.5**	**6.6**	**2.7**	**2.8**	**0.7**
Retired	0.0	0.0	0.5	0.6	60.4	62.2
Retired plus job	0.0	0.0	0.0	0.0	0.4	0.3
Family care	0.1	9.8	0.7	18.9	0.2	16.5
Family care plus job	0.0	0.3	0.0	0.8	0.0	0.2
Long-term sick/disabled	0.6	0.4	3.8	3.2	7.6	3.0
Long-term sick/dis. plus job	0.0	0.0	0.1	0.2	0.1	0.0
Something else	0.3	0.2	0.1	0.2	0.0	0.1
Something else plus job	0.1	0.2	0.0	0.1	0.1	0.0
Total Inactive	**1.1**	**10.9**	**5.3**	**23.9**	**68.7**	**82.3**
Full-time education	21.8	20.2	0.7	0.9	0.1	0.0
Full-time education plus job	11.1	14.5	0.3	0.3	0.0	0.0
Total Full-Time Education	**32.9**	**34.7**	**1.0**	**1.2**	**0.1**	**0.0**
Number of Observations	6,562	6,804	21,143	24,260	10,228	13,200

Notes: Data are from the first nine waves of the BHPS. Paid employment includes employees and excludes the self-employed. The main activity is the response to the question of "What describes your current situation" while the "plus job" corresponds to information from the questions "Did you do any paid work last week ?" and "Did you have a job you were away from last week ?". Maternity leave and government training are included as paid employment as most respondents in these categories also held jobs.

Table 2: Overlap Matching Summary

Percentage of Overlaps in Each Group	Young		Prime		Old		All
	Men	Women	Men	Women	Men	Women	
1 spell in both reports:							
- states match	83.98	83.92	93.46	91.21	95.95	97.20	92.33
- states do not match	6.63	7.23	3.36	4.79	3.39	2.34	4.15
2 or more spells in both reports:							
- states match	2.15	2.06	0.89	1.10	0.21	0.09	0.91
- states do not match	0.89	0.64	0.33	0.34	0.06	0.01	0.31
Loss from 1st report:							
1 spell less in 2nd	4.54	3.78	1.18	1.69	0.23	0.28	1.50
≥ 2 spells less in 2nd	0.39	0.21	0.08	0.10	0.00	0.00	0.09
Gain from 1st report:							
1 spell more in 2nd	1.35	2.06	0.67	0.76	0.15	0.08	0.69
≥ 2 spells more in 2nd	0.07	0.09	0.02	0.01	0.00	0.01	0.02
Number of Overlaps	4362	4659	16328	18590	7928	10101	61968

Table 3a: Spell Matching: All Types

Percentage of spells in initial report matched, redefined and disappeared	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
Employment spells:						
- matched	91.4	91.9	97.6	96.3	95.4	94.8
- redefined	4.0	4.4	1.5	2.2	4.0	4.4
- disappeared	4.6	3.6	0.9	1.6	0.6	0.8
Number of spells	2607	2458	12,075	12,920	1545	1564
Self-employment spells:						
- matched	71.6	78.9	93.9	84.3	90.7	80.0
- redefined	21.3	13.5	5.3	13.9	9.0	19.1
- disappeared	7.1	7.7	0.8	1.8	0.3	0.9
Number of spells	155	52	2629	1070	719	215
Unemployment spells:						
- matched	67.1	52.1	74.1	51.9	66.4	40.0
- redefined	19.2	30.0	19.2	38.2	32.3	57.3
- disappeared	13.7	17.9	6.7	9.9	1.3	2.7
Number of spells	562	363	1092	526	229	75
Out of labour force spells:						
- matched	51.6	86.1	87.4	90.8	98.5	98.9
- redefined	30.7	10.0	11.1	7.7	1.4	1.0
- disappeared	17.7	4.0	1.5	1.6	0.1	0.1
Number of spells	62	603	846	4509	5469	8284
Full-time education spells:						
- matched	91.4	91.5	88.2	84.4	83.3	0.0
- redefined	7.4	7.5	8.9	11.0	16.7	100.0
- disappeared	1.2	1.0	3.0	4.6	0.0	0.0
Number of spells	1387	1541	135	218	6	4

Table 3b: Spell Matching: Employment

Percentage of employment spells in first report	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
Perfect match	42.9	51.4	43.8	47.5	39.0	42.3
Shifted match	0.4	0.3	0.1	0.1	0.0	0.0
Lengthened match	24.0	19.6	24.9	23.0	28.9	26.2
Shortened match	24.2	20.6	28.8	25.8	27.6	26.3
Redefined as self-employment	0.5	0.1	0.7	0.6	1.1	1.1
Redefined as unemployment	1.2	1.1	0.5	0.3	0.2	0.1
Redefined as inactive	0.4	1.1	0.3	1.3	2.7	3.1
Redefined as full-time education	1.8	2.0	0.0	0.1	0.0	0.0
Disappeared into employment	1.8	1.5	0.5	0.9	0.2	0.4
Disappeared into self-employment	0.3	0.1	0.1	0.1	0.1	0.0
Disappeared into unemployment	1.3	0.7	0.3	0.1	0.1	0.1
Disappeared into inactive	0.0	0.5	0.0	0.4	0.3	0.4
Disappeared into full-time education	1.2	0.8	0.0	0.1	0.0	0.0
Change in spell length for matched spells (in months):						
mean	-1.0***	-1.1***	-1.0**	-1.4***	4.7**	1.1
(standard deviation)	(15.0)	(13.0)	(47.5)	(37.1)	(81.0)	(56.0)

Notes: A very small number of spells were redefined as or disappeared into two or more different states. The means are significantly different from zero at 1% (***), 5% (**) and 10% (*) level.

Table 3c: Spell Matching: Self-Employment

Percentage of self-employment spells in first report	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
Perfect match	31.6	25.0	37.8	35.0	32.8	28.4
Lengthened match	25.8	26.9	27.2	26.2	28.5	26.1
Shortened match	14.2	26.9	28.9	23.2	29.4	25.6
Redefined as employment	17.4	9.6	3.9	8.1	3.9	9.8
Redefined as unemployment	1.9	3.9	0.9	0.7	0.3	0.0
Redefined as inactive	1.3	0.0	0.5	5.0	4.9	9.3
Redefined as full-time education	0.0	0.0	0.0	0.2	0.0	0.0
Redefined as 2 different states	0.7	0.0	0.1	0.0	0.0	0.0
Disappeared into employment	4.5	1.9	0.4	0.7	0.1	0.5
Disappeared into unemployment	2.6	1.9	0.3	0.2	0.0	0.0
Disappeared into inactive	0.0	0.0	0.0	0.7	0.1	0.5
Disappeared into full-time education	0.0	3.9	0.0	0.2	0.0	0.0
Change in spell length for matched spells (in months):						
mean	1.3	-0.4	0.1	2.3	2.2	1.6
(standard deviation)	(21.0)	(8.2)	(48.8)	(42.4)	(88.1)	(75.8)

Notes: A very small number of spells were a shifted match or disappeared into two or more different states.

Table 3d: Spell Matching: Unemployment

Percentage of unemployment spells in first report	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
Perfect match	27.1	21.8	30.0	21.9	35.8	16.0
Shifted match	0.7	0.8	0.5	0.6	0.0	0.0
Lengthened match	19.4	16.5	21.1	12.9	15.7	6.7
Shortened match	19.9	13.0	22.6	16.5	14.9	17.3
Redefined as employment	11.2	9.6	6.4	5.7	1.8	6.7
Redefined as self-employment	1.6	0.6	3.6	1.0	1.8	2.7
Redefined as inactive	2.5	15.2	8.6	29.9	28.8	46.7
Redefined as full-time education	3.9	3.9	0.4	1.3	0.0	1.3
Redefined as 2 different states	0.0	0.8	0.3	0.4	0.0	0.0
Disappeared into employment	10.7	12.4	5.3	8.0	0.9	2.7
Disappeared into self-employment	0.2	0.3	0.7	0.2	0.0	0.0
Disappeared into unemployment	0.5	0.0	0.2	0.2	0.0	0.0
Disappeared into inactive	0.2	1.7	0.2	1.3	0.4	0.0
Disappeared into full-time education	2.1	3.0	0.3	0.2	0.0	0.0
Disappeared into 2 different states	0.0	0.6	0.0	0.0	0.0	0.0
Change in spell length for matched spells (in months):						
mean	0.4	-1.0	-0.3	-1.9	-0.7	-11.8
(standard deviation)	(11.1)	(18.8)	(24.1)	(28.0)	(28.0)	(41.0)

Table 3e: Spell Matching: Out of the Labour Force

Percentage of spells out of the labour force in first report	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
Perfect match	22.6	27.4	33.8	33.6	49.4	39.0
Shifted match	0.0	0.0	0.0	0.1	0.0	0.0
Lengthened match	14.5	28.2	27.2	28.5	24.9	29.2
Shortened match	14.5	30.5	26.4	28.7	24.3	30.6
Redefined as employment	8.1	4.5	2.5	3.3	0.5	0.5
Redefined as self-employment	1.6	0.0	1.2	1.0	0.3	0.3
Redefined as unemployment	16.1	5.0	6.9	3.1	0.5	0.3
Redefined as full-time education	4.8	0.3	0.6	0.2	0.0	0.0
Disappeared into employment	8.1	1.7	0.5	1.1	0.0	0.1
Disappeared into self-employment	0.0	0.0	0.4	0.1	0.0	0.0
Disappeared into unemployment	6.5	0.7	0.7	0.2	0.0	0.0
Disappeared into inactive	0.0	0.2	0.0	0.0	0.0	0.0
Disappeared into full-time education	3.2	1.3	0.0	0.1	0.0	0.0
Change in spell length for matched spells (in months):						
mean	-12.9	-0.2	0.8	1.8**	0.4	-1.5
(standard deviation)	(47.7)	(19.6)	(26.8)	(50.6)	(53.6)	(100.7)

Notes: A very small number of spells were a shifted match or were redefined as or disappeared into two or more different states. The means are significantly different from zero at 1% (***), 5% (**) and 10% (*) level.

Table 3f: Spell Matching: Full-time Education

Percentage of full-time education spells in first report	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
Perfect match	27.4	28.4	45.9	53.2	66.7	0.0
Lengthened match	27.7	26.7	16.3	12.8	0.0	0.0
Shortened match	36.3	36.5	25.9	18.4	16.7	0.0
Redefined as employment	4.9	5.7	3.7	4.6	0.0	0.0
Redefined as self-employment	0.1	0.1	0.0	0.0	0.0	0.0
Redefined as unemployment	1.7	0.8	3.0	1.8	0.0	0.0
Redefined as inactive	0.5	0.8	2.2	4.1	16.7	100.0
Redefined as 2 different states	0.1	0.1	0.0	0.5	0.0	0.0
Disappeared into employment	0.9	0.8	2.2	1.8	0.0	0.0
Disappeared into self-employment	0.1	0.0	0.0	0.0	0.0	0.0
Disappeared into unemployment	0.1	0.1	0.7	0.0	0.0	0.0
Disappeared into inactive	0.1	0.1	0.0	2.8	0.0	0.0
Change in spell length for matched spells (in months):						
mean	-9.7***	-13.3***	-1.3	-4.1	-2.4	n/a
(standard deviation)	(77.4)	(81.3)	(37.4)	(52.9)	(5.4)	

Notes: The means are significantly different from zero at 1% (***), 5% (**) and 10% (*) level.

Table 4: Multinomial Logit Models for Spell Match By Gender: Effects of Spell Length

	Men				Women			
	Probability Redefined		Probability Disappeared		Probability Redefined		Probability Disappeared	
Regressors	RRR	std. err.	RRR	std. err.	RRR	std. err.	RRR	std. err.
Type of spell:								
employment	omitted	omitted	omitted	omitted	omitted	omitted	omitted	omitted
self-employment	3.623***	0.433	0.395***	0.106	6.976***	0.858	4.518***	1.609
unemployment	8.532***	0.859	2.678***	0.438	16.977***	1.763	7.026***	1.349
out of lab. force	3.181***	0.426	1.516	0.489	2.948***	0.274	0.480***	0.079
education	2.511***	0.506	0.339***	0.101	3.214***	0.533	0.523**	0.135
Length of spell in months:								
employment	0.992***	0.001	0.872***	0.009	0.997***	0.001	0.903***	0.006
self-employment	0.995***	0.001	0.959***	0.009	0.997***	0.001	0.777***	0.044
unemployment	1.002	0.001	0.860***	0.018	1.009***	0.002	0.797***	0.029
out of lab. force	0.983***	0.002	0.817***	0.032	0.990***	0.001	0.969***	0.004
education	1.000	0.001	0.983***	0.004	0.998	0.001	0.976***	0.006
Overlap length	0.923	0.717	1.984***	0.254	0.983	0.070	2.124***	0.256
(Overlap length)^2	1.012	0.012	0.942***	0.017	1.002	0.011	0.926***	0.016
Pseudo R^2	0.178				0.177			
# of observations	29,518				34,402			

Notes: RRR denotes the relative risk ratio and std. err. denotes the standard error. The omitted outcome category is the probability that the spell matched. Ratios are significantly different from one at 1% (***), 5% (**) and 10% (*) level. Chi^2 tests were used to test whether the ratios were significantly different from each other. For all spell type and length variables, the ratios for the probability of redefinition were significantly different at the 1 percent level from those for the probability of disappearance except for out of labour force spells for men (which are significantly different at the 5 percent level) and spells of self-employment for women (which are not significantly different). For men, the ratios for the probability of redefinition are significantly different at the 1 percent level for self-employment and unemployment spell types, for unemployment and out of the labour force spell types, for out of the labour force and education spell types and between all the spell length variables except between employment and self-employment and between unemployment and education (which are not significantly different). The ratios are significantly different at the 10 percent level for self-employment and education spell types. For women, the ratios for the probability of redefinition are significantly different at the 1 percent level between all spell type variables except between out of the labour force and education (which are not significantly different). The ratios for the probability of redefinition are significantly different at the 1 percent level between all spell length variables except between employment and self-employment, between employment and education and between self-employment and education (which are not significantly different). For men, the ratios for the probability of disappearance are significantly different at the 1 percent level between all spell type variables except between self-employment and education (which are not significantly different) and between unemployment and out of the labour force (which are significantly different at the 10 percent level). The ratios for the probability of disappearance are significantly different at the 1 percent level between all spell length variables except between employment and self-employment and between unemployment and education (which are not significantly different). For women, the ratios for the probability of disappearance are significantly different at the 1 percent level between all spell type variables except between self-employment and unemployment and between out of the labour force and education (which are not significantly different). The ratios for the probability of disappearance are significantly different at the 1 percent level between all spell length variables except between self-employment and unemployment and between out of the labour force and education (which are not significantly different).

Table 5a: Predicted Proportions of Spells Redefined and Disappeared At Different Spell Lengths: Men

Spell Type:	Percentage of spells redefined at spell length:			Percentage of spells disappeared at spell length:		
	3 months	12 months	60 months	3 months	12 months	60 months
Employment	2.8	2.8	2.0	8.1	2.5	0.0
Self-employment	9.9	9.6	7.9	4.1	2.9	0.4
Unemployment	18.4	21.0	23.4	15.5	4.5	0.0
Out of labour force	8.1	7.6	3.6	9.4	1.7	0.0
Full-time education	7.2	7.3	7.4	3.9	3.4	1.6

Notes: The percentages are predicted using the model shown in table 4, with the overlap length set to the median value of 2 months.

Table 5b: Predicted Proportions of Spells Redefined and Disappeared At Different Spell Lengths: Women

Spell Type:	Percentage of spells redefined at spell length:			Percentage of spells disappeared at spell length:		
	3 months	12 months	60 months	3 months	12 months	60 months
Employment	3.0	3.0	2.7	7.9	3.3	0.0
Self-employment	15.6	18.0	16.0	17.0	2.1	0.0
Unemployment	29.1	37.6	49.6	21.2	3.2	0.0
Out of labour force	8.4	7.8	5.2	4.5	3.5	0.8
Full-time education	9.2	9.2	8.7	5.0	4.1	1.3

Notes: The percentages are predicted using the model shown in table 4, with the overlap length set to the median value of 2 months.

Table 6: Predicted Proportions of Spells Matched Controlling for Spell and Overlap Length: Effects of Spell Type and Gender and Age

Predicted Percentage of Spells Matched	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
Employment	93.9	94.4	97.7	96.7	94.2	93.1
sig. diff. from age:	prime ***	prime ***	old ***	old ***	young -	young -
sig. diff. by gender:	-		***		-	
Self-employment	78.8	84.6	93.0	83.6	85.5	70.8
sig. diff. from age:	prime ***	prime -	old ***	old ***	young ***	young ***
sig. diff. by gender:	-		***		***	
Unemployment	76.8	63.7	80.4	60.3	67.1	40.6
sig. diff. from age:	prime *	prime -	old ***	old ***	young ***	young ***
sig. diff. by gender:	***		***		***	
Out of labour force	59.6	88.9	87.9	89.8	97.9	98.1
sig. diff. from age:	prime ***	prime -	old ***	old ***	young ***	young ***
sig. diff. by gender:	***		*		-	
Education	87.7	87.7	91.4	88.9	n/a	n/a
sig. diff. from age:	prime -	prime -	n/a	n/a	n/a	n/a
sig. diff. by gender:	-		-		n/a	
Sig. diff. by spell type:						
- emp & self-emp	***	***	***	***	***	***
- emp & unemp	***	***	***	***	***	***
- emp & out of LF	***	***	***	***	***	***
- emp & educ	***	***	***	***	n/a	n/a
- self-emp & unemp	-	***	***	***	***	***
- self-emp & out of LF	***	-	***	***	***	***
- self-emp & educ	***	-	-	***	n/a	n/a
- unemp & out of LF	***	***	***	***	***	***
- unemp & educ	***	***	***	***	n/a	n/a
- out of LF & educ	***	-	-	-	n/a	n/a

Notes: The percentages are significantly different across categories at 1% (***), 5% (**) and 10% (*) levels, while the symbol "-" denotes no significant difference in the percentage. The percentages of spells that are matched are estimated from a logit model for the probability of a match that included dummy variables for each category shown and control variables for the length of spell and overlap length (linear and squared terms). The percentages are predicted at the median spell length of 52 months and median overlap length of 2 months. The spell type full-time education was omitted for old men and women due to a small number of observations in these categories (6 and 4 respectively). The symbol "n/a" denotes not applicable for this missing category.

Table 7: Distribution of Time Use in Wave Overlaps

Mean percentage of time in:	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
Employment:						
Initial report	54.4	48.4	72.4	67.4	19.2	15.3
Later report	56.3	49.6	72.6	67.4	19.2	15.2
Ratio of later/initial	1.035 ***	1.025 ***	1.003	1.000	1.000	0.993
Self-Employment:						
Initial report	3.3	1.0	15.9	5.6	9.0	2.1
Later report	3.0	1.0	15.8	5.5	8.7	2.1
Ratio of later/initial	0.909 **	1.000	0.994	0.982	0.967 ***	1.000
Unemployment:						
Initial report	10.3	6.1	5.9	2.4	2.8	0.7
Later report	9.2	5.0	5.5	2.2	2.3	0.5
Ratio of later/initial	0.893 ***	0.820 ***	0.932 ***	0.917 *	0.821 ***	0.714 **
Out of Labour Force:						
Initial report	1.1	12.3	5.0	23.6	68.9	81.9
Later report	1.4	13.0	5.4	23.8	69.7	82.1
Ratio of later/initial	1.273 **	1.057 ***	1.080 ***	1.008	1.012 ***	1.002
Education:						
Initial report	30.9	32.2	0.8	1.0	0.0	0.0
Later report	30.2	31.3	0.8	1.1	0.0	0.0
Ratio of later/initial	0.977 **	0.972 ***	1.000	1.100	n/a	n/a

Notes: The ratios are significantly different from one at 10%(*) 5%(**) and 1%(***) level using t-test for paired observations between the initial and later reports.

Table 8: Numbers of Spells in Wave Overlaps

Total number of spells of:	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
All types:						
Initial report	4773	5017	16777	19243	7968	10143
Later report	4604	4923	16675	19037	7962	10125
Ratio of later/initial	0.965 ***	0.981 ***	0.994 ***	0.989 ***	0.999	0.998 ***
Employment:						
Initial report	2607	2458	12075	12920	1545	1564
Later report	2590	2474	12050	12789	1537	1551
Ratio of later/initial	0.993	1.007	0.998	0.990 ***	0.995	0.992
Self-Employment:						
Initial report	155	52	2629	1070	719	215
Later report	135	48	2609	1037	690	214
Ratio of later/initial	0.871 **	0.923	0.992	0.969 *	0.960 ***	0.995
Unemployment:						
Initial report	562	363	1092	526	229	76
Later report	460	277	973	478	192	53
Ratio of later/initial	0.819 ***	0.763 ***	0.891 ***	0.909 **	0.838 ***	0.697 ***
Out of Labour Force:						
Initial report	62	603	846	4509	5469	8284
Later report	77	632	908	4509	5537	8303
Ratio of later/initial	1.242 *	1.048 **	1.073 ***	1.000	1.012 **	1.002
Education:						
Initial report	1387	1541	135	218	6	4
Later report	1342	1492	135	224	6	4
Ratio of later/initial	0.968 ***	0.968 ***	1.000	1.028	1.000	1.000

Notes: The ratios are significantly different from one at 10%(*) 5%(**) and 1%(***) level using t-test for paired observations between the initial and later reports.

Table 9: Division of Time Over Prior 8-24 Months

Percentage of Time in:	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
Employment						
A. Traditional Panel	86.2	76.0	91.1	75.5	33.2	20.5
B. Latest Inter. Rules	86.4	75.9	90.8	75.1	33.0	20.9
C. Reconciled	84.6	75.7	89.8	74.4	33.7	21.2
D. Selected No Probs.	89.3	80.0	93.1	78.3	33.1	19.2
Differences Between Paired Observations:						
Method A and B	86.5 86.8*	76.3 76.2	91.2 91.3***	75.6 75.5	33.2 33.0**	20.5 20.4
Method A and C	86.3 86.2	76.1 76.1	91.1 91.1	75.5 75.5	33.2 33.2	20.6 20.5
Method B and C	86.5 86.2**	76.3 76.3	90.9 90.8**	75.2 75.1	33.1 33.3***	20.8 20.8
Unemployment						
A. Traditional Panel	12.1	6.8	5.2	1.9	2.8	0.6
B. Latest Inter. Rules	11.6	6.5	5.1	2.1	2.5	0.6
C. Reconciled	13.6	7.1	5.9	2.4	3.0	0.8
D. Selected No Probs.	9.7	5.3	4.1	1.2	2.1	0.2
Differences Between Paired Observations:						
Method A and B	11.8 11.3***	6.7 6.2***	5.2 5.0***	1.9 1.9	2.8 2.5***	0.6 0.5**
Method A and C	12.1 12.1	6.8 6.8	5.2 5.2	1.9 2.0	2.8 2.8	0.6 0.7
Method B and C	11.5 12.0***	6.2 6.8***	5.1 5.3***	2.1 2.2	2.5 2.9***	0.6 0.7**
Out of the Labour Force						
A. Traditional Panel	1.7	17.2	3.7	22.5	64.0	78.8
B. Latest Inter. Rules	2.0	17.6	4.1	22.7	64.4	78.4
C. Reconciled	1.8	17.2	4.3	23.2	63.3	78.0
D. Selected No Probs.	1.0	14.7	2.8	20.5	64.9	80.6
Differences Between Paired Observations:						
Method A and B	1.7 1.9	16.9 17.5***	3.6 3.7*	22.4 22.6	64.0 64.5***	78.9 79.1**
Method A and C	1.7 1.7	17.1 17.1	3.7 3.6	22.5 22.5	64.0 64.0	78.8 78.8
Method B and C	2.0 1.8	17.5 16.9***	4.1 3.9***	22.7 22.7	64.4 63.8***	78.6 78.5

Notes: Average time in a given state is significantly different between methods at 10%(*) 5%(**) and 1%(***) level using t-tests for paired observations. Employment includes self-employment. The numbers of valid observations (for the unpaired means) for methods A to D are 2302, 2394, 2698, and 1984 for young men; 2523, 2637, 2871, and 2186 for young women; 17632, 17878, 18955 and 16461 for prime men; 19255, 19782, 21201 and 17253 for prime women; 8819, 8989, 9365 and 8067 for old men; and 10951, 11164, 11673, and 10265 for old women.

Table 10a: Monthly Transition Rates Over Prior 8-24 Months: From Employment

Percentage of Employed Who Move in Subsequent Month:	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
To Another Employer						
A. Traditional Panel	1.90	2.19	0.83	1.03	0.30	0.35
B. Latest Interview Rules	1.47	1.74	0.64	0.80	0.19	0.21
C. Reconciled	1.97	2.07	0.84	1.03	0.42	0.37
D. Selected No Problems	1.59	1.94	0.70	0.94	0.26	0.27
Significance of differences:						
Method A and B	***	***	***	***	***	***
Method A and C	**	***			***	
Method A and D	***	***	***	***	***	***
Method B and C	***	***	***	***	***	***
Method B and D	***	***	***	***	***	***
Method C and D	***	***	***	***	***	***
To Unemployment						
A. Traditional Panel	1.74	1.06	0.52	0.33	0.36	0.21
B. Latest Interview Rules	1.29	0.75	0.40	0.24	0.26	0.14
C. Reconciled	1.86	1.05	0.55	0.34	0.37	0.23
D. Selected No Problems	1.09	0.73	0.36	0.22	0.27	0.12
Significance of differences:						
Method A and B	***	***	***	***	***	***
Method A and C	**		***			
Method A and D	***	***	***	***	***	***
Method B and C	***	***	***	***	***	***
Method B and D	***		***	***		**
Method C and D	***	***	***	***	***	***
Out of the Labour Force						
A. Traditional Panel	0.13	0.62	0.08	0.52	0.97	1.30
B. Latest Interview Rules	0.10	0.48	0.07	0.35	0.75	0.94
C. Reconciled	0.15	0.62	0.11	0.58	1.08	1.43
D. Selected No Problems	0.08	0.43	0.06	0.30	0.70	0.85
Significance of differences:						
Method A and B	***	***	***	***	***	***
Method A and C	*		***	***	***	***
Method A and D	***	***	***	***	***	***
Method B and C	***	***	***	***	***	***
Method B and D	**	**	***	***	*	***
Method C and D	***	***	***	***	***	***

Notes: Average transition is significantly different between methods at 10%(*) 5%(**) and 1%(***) level. A spell of self-employment is counted as a different employer (oneself). The numbers of monthly observations for methods A to D are 32509, 33361, 36569, and 29015 for young men; 31060, 32187, 34756 and 28317 for young women; 257969, 260286, 272649 and 245919 for prime men; 234172, 238535, 252729 and 217229 for prime women; 46930, 47513, 50490 and 42707 for old men; 36373, 37654, 39564 and 31861 for old women.

Table 10b: Monthly Transition Rates Over Prior 8-24 Months: From Unemployment

Percentage of Unemployed Who Move in Subsequent Month:	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
To Employment						
A. Traditional Panel	10.84	9.91	6.34	10.28	2.83	4.89
B. Latest Interview Rules	7.98	8.65	5.32	7.71	2.19	3.80
C. Reconciled	10.47	10.34	6.10	9.16	2.55	4.57
D. Selected No Problems	8.87	9.70	5.93	11.47	2.62	6.85
Significance of differences:						
Method A and B	***	***	***	***	***	**
Method A and C				***		
Method A and D	***		**	***		**
Method B and C	***	***	***	***	*	
Method B and D	**	**	***	***	*	***
Method C and D	***			***		**
Out of the Labour Force						
A. Traditional Panel	0.47	2.42	1.02	2.76	3.06	4.80
B. Latest Interview Rules	0.24	0.91	0.38	0.58	0.82	1.04
C. Reconciled	0.51	2.49	1.18	2.76	3.16	4.63
D. Selected No Problems	0.09	1.18	0.35	0.50	1.18	1.10
Significance of differences:						
Method A and B	***	***	***	***	***	***
Method A and C			**			
Method A and D	***	***	***	***	***	***
Method B and C	***	***	***	***	***	***
Method B and D	***	*			***	
Method C and D	***	***	***	***	***	***

Notes: Average transition is significantly different between methods at 10%(*) 5%(**) and 1%(***) level. Employment includes self-employment. The numbers of monthly observations for methods A to D are 4639, 4536, 5877 and 3168 for young men; 2774, 2739, 3253 and 1866 for young women; 15026, 14747, 17845 and 11055 for prime age men; 6165, 6925, 8120 and 3373 for prime age women; 3923, 3661, 4520 and 2707 for old men; and 1146, 1157, 1511 and 365 for old women.

Table 10c: Monthly Transition Rates Over Prior 8-24 Months: From Out of the Labour Force

Percentage of Out of Labour Force Who Move in Subsequent Month:	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
To Employment						
A. Traditional Panel	5.97	1.78	0.95	1.52	0.10	0.10
B. Latest Interview Rules	4.55	1.13	0.62	1.08	0.04	0.04
C. Reconciled	5.73	1.84	0.94	1.60	0.10	0.11
D. Selected No Problems	7.65	1.26	0.71	1.00	0.03	0.04
Significance of differences:						
Method A and B	**	***	***	***	***	***
Method A and C				***		*
Method A and D		***	***	***	***	***
Method B and C	*	***	***	***	***	***
Method B and D	***			***	***	
Method C and D	*	***	***	***	***	***
To Unemployment						
A. Traditional Panel	3.30	0.52	1.03	0.24	0.06	0.02
B. Latest Interview Rules	1.52	0.23	0.34	0.09	0.01	0.00
C. Reconciled	3.05	0.56	0.99	0.30	0.07	0.02
D. Selected No Problems	2.14	0.25	0.37	0.04	0.01	0.00
Significance of differences:						
Method A and B	***	***	***	***	***	***
Method A and C				***	*	***
Method A and D	*	***	***	***	***	***
Method B and C	***	***	***	***	***	***
Method B and D				***		
Method C and D		***	***	***	***	***

Notes: Average transition is significantly different between methods at 10%(*) 5%(**) and 1%(***) level. Employment includes self-employment. The numbers of monthly observations for methods A to D are 637, 791, 786 and 327 for young men; 6955, 7442, 7928 and 5141 for young women; 10461, 11847, 13038 and 7469 for prime age men; 69798, 72124, 78698 and 56997 for prime age women; 90842, 93002, 94986 and 84074 for old men; and 139148, 140821, 145878 and 133374 for old women.

Table 11: Survival Models

Estimated Median Survival Time from A Weibull Model	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
Employment (Tenure)						
A. Traditional Panel	10.6	12.1	31.8	26.1	61.2	50.4
B. Latest Interview Rules	15.5	16.7	42.3	33.9	68.1	54.7
C. Reconciled	11.2	13.2	33.4	27.4	62.0	51.0
D. Selected No Problems	11.7	13.3	34.0	27.5	63.1	51.4
Significance of differences:						
Method A and B	***	***	***	***	***	***
Method A and C		*	**	***		
Method A and D	**	*	***	***		
Method B and C	***	***	***	***	***	***
Method B and D	***	***	***	***	***	***
Method C and D						
Unemployment						
A. Traditional Panel	4.4	4.1	5.8	4.7	8.5	8.4
B. Latest Interview Rules	6.0	5.3	7.4	6.4	9.7	10.6
C. Reconciled	4.9	4.5	6.2	5.2	9.3	8.6
D. Selected No Problems	4.7	4.1	6.0	4.4	8.2	8.3
Significance of differences:						
Method A and B	***	***	***	***		*
Method A and C	*		**	**		
Method A and D						
Method B and C	***	**	***	***		*
Method B and D	***	***	***	***	*	*
Method C and D				***		
Out of the Labour Force						
A. Traditional Panel	4.2	15.6	14.1	30.8	136.9	149.5
B. Latest Interview Rules	4.8	22.9	24.5	45.2	218.8	204.1
C. Reconciled	5.4	17.8	16.3	32.8	145.9	157.3
D. Selected No Problems	3.7	18.8	16.2	35.7	161.5	161.4
Significance of differences:						
Method A and B		**	***	***	***	***
Method A and C				*		
Method A and D				***	**	*
Method B and C			***	***	***	***
Method B and D			***	***	***	***
Method C and D	**			**		

Notes: Employment includes self-employment and each employment spell covers time with a given employer. Weibull models that included dummy variables for each method were estimated for each of the three states and six groups. The significance of the differences was calculated as the significance of the differences in the coefficients on the respective dummy variables at the 10%(*) 5%(**) and 1%(***) level. The number of spells (observations) in the model for employment spells was 16,178 for young men, 14,395 for young women, 98,563 for prime aged men, 110,134 for prime aged women, 55,012 for old men and 59,988 for old women. The number of spells (observations) in the model for unemployment spells was 6,213 for young men, 3,947 for young women, 18,603 for prime aged men, 11,128 for prime aged women, 4,019 for old men and 1,956 for old women. The number of spells

(observations) in the model for spells out of the labour force was 1,013 for young men, 1,979 for young women, 4,138 for prime aged men, 22,205 for prime aged women, 13,393 for old men and 22,859 for old women. The differences in the equality of the survival functions between the methods were also tested using a log-rank test and a Wilcoxon (Breslow) test, with a few contrasting results to the significance of the differences between the coefficients in the Weibull model. In particular, for employment spells, the log rank test found no significant difference between methods A and C for young women, while the Wilcoxon test found significant differences between methods A and D for old men and women and between methods C and D for prime men. For unemployment spells, both tests found no significant differences between methods A and C for young men, between B and C for old women and between B and D for old men. For spells out of the labour force, the Wilcoxon test found significant differences between methods A and B and B and D for young men and between C and D for old men, while both tests found no significant difference between methods C and D for young men.

Table 12: Average Years of Employment Experience

Average Years of Employment Experience	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
Method:						
A. Traditional Panel	1.88	1.77	20.71	15.93	42.40	28.02
B. Latest Interview Rules	1.97	1.83	20.72	15.85	41.62	26.90
C. Reconciled	1.99	1.93	20.50	15.76	42.43	28.13
D. Selected No Problems	1.73	1.66	21.11	16.17	42.29	27.70
Differences Between Paired Observations:						
Method A and B	1.88 1.93***	1.78 1.79	20.73 20.78***	15.96 15.89***	42.40 41.66***	28.02 26.85***
Method A and C	1.89 1.92***	1.78 1.83***	20.74 20.73*	15.93 15.90***	42.39 42.40	28.02 27.96***
Method B and C	1.93 1.91***	1.81 1.83*	20.83 20.77***	15.90 15.94***	41.66 42.42***	26.93 28.02***

Notes: Employment includes employment and self-employment. The samples include all individual regardless of current employment status. Average number of years is significantly different between methods at 10%(*) 5%(**) and 1%(***) level using t-tests for paired observations. The numbers of valid observations (for the unpaired means) for methods A to D are 4918, 5008, 5445 and 4170 for young men; 5198, 5243, 5669 and 4520 for young women; 13951, 14653, 15242 and 12435 for prime men; 16977, 17872, 18813 and 14435 for prime women; 8139, 8453, 8651 and 7450 for old men; 10894, 11214, 11415 and 10072 for old women.

Table 13: Average Years of Employer Tenure for Those Currently in Paid Employment

Average Years of Employer Tenure	Young		Prime		Old	
	Men	Women	Men	Women	Men	Women
Method:						
A. Traditional Panel	1.64	1.57	7.31	5.50	13.22	10.98
B. Latest Interview Rules	1.86	1.78	8.50	6.56	16.17	14.17
C. Reconciled	1.70	1.68	7.49	5.66	13.39	11.32
D. Selected No Problems	1.63	1.58	7.40	5.59	13.99	11.83
Differences Between Paired Observations:						
Method A and B	1.63 1.82***	1.57 1.73***	7.36 8.38***	5.53 6.50***	13.38 15.95***	11.09 14.11***
Method A and C	1.64 1.70***	1.57 1.66***	7.31 7.59***	5.50 5.72***	13.23 13.56***	10.99 11.57***
Method B and C	1.82 1.69***	1.73 1.66***	8.39 7.61***	6.52 5.74***	15.96 13.62***	14.13 11.58***

Notes: The samples include only those currently in paid employment and excludes those in self-employment. Average number of years is significantly different between methods at 10%(*) 5%(**) and 1%(***) level using t-tests for paired observations. The numbers of valid observations (for the unpaired means) for methods A to D are 2874, 2849, 3007, and 2774 for young men; 2730, 2737, 2861 and 2650 for young women; 11602, 11698, 11997 and 11419 for prime men; 13247, 13264, 13770 and 12902 for prime women; 1595, 1598, 1661 and 1491 for old men; and 1688, 1686, 1755 and 1536 for old women.

Table 14a: Wage Returns to Tenure and Experience for those in Paid Employment: Men

Dependent variable: ln(wage)	Traditional Panel (Method A)		Latest Interview Rules (Method B)		Reconciled (Method C)		Selected No Problems (Method D)	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Base model								
tenure	0.018***	0.002	0.014***	0.001	0.019***	0.001	0.015***	0.002
tenure^2	-0.000***	0.000	-0.000***	0.000	-0.000***	0.000	-0.000***	0.000
experience	0.054***	0.001	0.053***	0.001	0.055***	0.001	0.055***	0.001
experience^2	-0.001***	0.000	-0.001***	0.000	-0.001***	0.000	-0.001***	0.000
constant	1.553***	0.011	1.561***	0.011	1.518***	0.011	1.556***	0.012
Adjusted R^2	0.212		0.203		0.222		0.207	
# of obs.	12086		12579		13214		10857	
Add interactions for method B:								
tenure	-0.003	0.002						
tenure^2	0.000	0.000						
experience	-0.001	0.002						
experience^2	0.000	0.000						
constant	0.009	0.016						
Joint Sig.								
Add interactions for method C:								
tenure	0.001	0.002	0.005**	0.002				
tenure^2	-0.000	0.000	-0.000**	0.000				
experience	0.002	0.002	0.002	0.002				
experience^2	-0.000	0.000	-0.000	0.000				
constant	-0.036**	0.015	-0.044***	0.015				
Joint Sig.			**					
Add interactions for method D:								
tenure	-0.002	0.002	0.001	0.002	-0.004	0.002		
tenure^2	0.000	0.000	-0.000	0.000	0.000	0.000		
experience	0.002	0.002	0.002	0.002	-0.000	0.002		
experience^2	-0.001	0.000	-0.000*	0.000	-0.000	0.000		
constant	0.003	0.016	-0.005	0.016	0.039**	0.016		
Joint Sig.					**			

Notes: Paid employment does *not* include self-employment. Coeff. denotes estimated coefficient and s.e. denotes standard error. Coefficients are significantly different from zero at 10%(*) 5%(**) and 1%(***) significance levels. Joint sig. shows the joint significance of the interaction variables added to the basic model.

Table 14b: Wage Returns to Tenure and Experience for those in Paid Employment: Women

Dependent variable: ln(wage)	Traditional Panel (Method A)		Latest Interview Rules (Method B)		Reconciled (Method C)		Selected No Problems (Method D)	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Base model								
tenure	0.026***	0.002	0.020***	0.002	0.024***	0.002	0.022***	0.002
tenure^2	-0.001***	0.000	-0.000***	0.000	-0.000***	0.000	-0.000***	0.000
experience	0.024***	0.001	0.022***	0.001	0.027***	0.001	0.026***	0.002
experience^2	-0.001***	0.000	-0.001***	0.000	-0.001***	0.000	-0.001***	0.000
constant	1.535***	0.011	1.556***	0.011	1.509***	0.011	1.544***	0.013
Adjusted R^2	0.064		0.057		0.067		0.059	
# of obs.	13817		14370		15248		12061	
Add interactions for method B:								
tenure	-0.006**	0.002						
tenure^2	0.000**	0.000						
experience	-0.002	0.002						
experience^2	0.000	0.000						
constant	0.021	0.016						
Joint Sig.	**							
Add interactions for method C:								
tenure	-0.003	0.003	0.004	0.002				
tenure^2	0.000	0.000	-0.000	0.000				
experience	0.003	0.002	0.005**	0.002				
experience^2	-0.000	0.000	-0.000**	0.000				
constant	-0.026	0.016	-0.047***	0.016				
Joint Sig.			***					
Add interactions for method D:								
tenure	-0.004	0.003	0.002	0.003	-0.002	0.003		
tenure^2	0.000	0.000	-0.000	0.000	-0.000	0.000		
experience	0.002	0.002	0.004*	0.002	-0.001	0.002		
experience^2	-0.000	0.000	-0.000**	0.000	0.000	0.000		
constant	0.009	0.017	-0.012	0.017	0.035**	0.017		
Joint Sig.			***		**			

Notes: Paid employment does *not* include self-employment. Coeff. denotes estimated coefficient and s.e. denotes standard error. Coefficients are significantly different from zero at 10%(*) 5%(**) and 1%(***) significance levels. Joint sig. tests the joint significance of the interaction variables added to the basic model.