Disentangling Overlapping Seams: The experience of the HILDA Survey

Nicole Watson

Melbourne Institute of Applied Economic and Social Research

University of Melbourne

Acknowledgement: This paper uses the unit record data file from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Survey project was initiated and is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research. The findings and views reported in this paper, however, are those of the author and should not be attributed to either FaHCSIA or the Melbourne Institute.

Abstract

An issue unique to longitudinal surveys is seam effects. These occur when there is a tendency for changes in the data to unusually concentrate in adjoining periods from different interviews. One component of the Household, Income and Labour Dynamics in Australia (HILDA) Survey subject to seam effects is the labour market activity calendar. In this calendar respondents are asked to recall the various jobs they have had over a 14 to 18 month period, the time spent in unemployment and the time spent outside the labour force. As the calendar is administered every wave, an overlap of 2 to 6 months results, depending on when the respondent is interviewed.

In this paper, we separately model the likelihood that respondents will make three types of errors in the activity calendar. These errors are: i) reporting a spell in the first version of events and not in the second; ii) misplacing a spell in the second version of events compared to the first; and iii) reporting a spell in the second version of events and not in the first. The characteristics considered in the model include the various causes of errors in dating events, such as spell length, spell type, duration of the overlapping seam, recall ability of the respondent, and characteristics of the interview that may affect the respondent's recall. The overlapping seam also permits the study of measurement error over time to identify whether the same people continually make the same mistakes.

With a better understanding of the types of errors that respondents make, the HILDA team hopes to construct a consolidated labour market activity spell file that will encourage greater data use of a section of the HILDA Survey that has so far been underutilised.

1. Introduction

Longitudinal studies often incorporate questions that require the respondent to report activities over the intervening period between the current interview and the previous interview. Inconsistencies in their recall compared to the data from the previous interview gives rise to seam effects. These occur when there is a tendency for changes in the data to unusually concentrate in adjoining periods from different interviews (Tourangeau *et al.*, 2000). It is not uncommon for the size of transitions across adjoining periods (the seam) to be between two to eight times the size of the transitions that are not at the seam. Burkhead and Coder (1985), for example, found that the transitions in the quarterly US Survey of

Income and Program Participation for the receiving a particular benefit, Medicade, were eight times higher at the seam compared to the average within wave transitions. Within the same survey, Tourangeau *et al.* (2000) found that the employment status transitions at the seam were twice those off the seam, total family income changes were three times higher at the seam compared to off-seam months, and individual social security income changes were six times higher at the seam. Another example comes from Statistics Canada's annual Labour Market Activity Survey – Lemaître (1992) found that the movements into and out of self-employment were three to four times larger at the seam compared to the average within wave transitions.

Research on seam effects has focused on three areas. These are: i) understanding why seam effects occur (e.g., Tourangeau *et al.*, 2000, Jäckle, 2008a; Callegaro, 2008); ii) identifying techniques to handle seam effects in data analysis (e.g., Weidman, 1986; Halpin, 1998; Rips *et al.*, 2003; Lynn *et al.*, 2005); and iii) to reduce the likelihood of seam effects at the data collection stage (e.g., Jäckle and Lynn, 2007; Callegaro and Belli, 2007; Jäckle, 2008b; Moore *et al.*, 2009). Using data from the labour market activity calendar in the HILDA Survey, this paper seeks to contribute to the first of these three areas.

Following a summary of the literature on the factors affecting recall errors, this paper describes the HILDA calendar in detail. The calendar collects information from a fixed point in the previous year to the date of interview, resulting in two reports for the same portion of time (called the 'overlapping seam'). This overlapping seam gives us the opportunity to investigate a number of matching methods and study the recall errors respondents make. This will provide a sound basis for the construction of a consolidated labour market activity spell file for HILDA users. We hope that such a spell file will encourage use of a hereto underutilized section of the HILDA data (only three papers seem to have used the calendar data to date: Dockery, 2004; Carroll, 2006; Watson, 2008).

2. Factors Affecting Spell Recall Errors

The types of errors which can occur when a spell is being recorded during an interview are:

- Omissions – the respondent may forget to record a particular spell.
- Misclassification – the respondent may modify the type of activity recorded, perhaps a period of unemployment is misclassified to a period not in the labour force as their movement between these two spells was somewhat unclear to the respondent.

- Forward or backward telescoping – the respondent may bring forward events in time or they may push them back in time, especially around the beginning of the reference period.

- Estimation – when the respondent does not want to or cannot remember, they may extrapolate from recent events (termed 'constant wave response'), or they may resort to averaging or guessing.

- Misinterpretation of the question – there may be different interpretations of how a spell should be recorded, for example, self-employment with very irregular hours.

The factors that affect these spell recall errors can be divided into five categories: spell characteristics, the complexity of the recall; the respondent characteristics; interview characteristics and survey process characteristics. The findings from the literature for each of these areas are summarized below. (Both Jäckle (2008a) and Callegaro (2008) provide excellent descriptions of these factors for readers who want more.)

Spell characteristics

Shorter spells are more likely to be dropped or misplaced (Halpin, 1998; Paull, 2002; Jäckle, 2008a). Jäckle (2008a) shows that seam effects are smaller for events with clearly defined dates (such as transitions to full-time education or employment) than fuzzy dates (such as transition to self-employment or not in the labour force).

Complexity of the recall

The difficulty of the recall task and the accuracy of the inference strategies that respondents use when they cannot remember exactly affect the size of the seam effect (Rips *et al.*, 2003). In support of this, Callegaro and Belli (2007), using US Panel Study of Income Dynamics (PSID), found that people with complicated job histories were more likely to give inconsistent data. In contract, Jäckle (2008a) failed to find that the higher the number of spells in the recall period (based on administrative data) affected the size of the seam effect.

We would expect that as the length of the reference period increases, the size of the seam effect would also increase. Paull (2002) found evidence of this using the British Household Panel Survey (BHPS) data, though this effect was absent in the PSID data (Hill, 1987).

Respondent characteristics

Forgetting seemed to be the main cause of inconsistencies over time in labour market activities (Jäckle and Lynn, 2007) and in changes in income (Kalton and Miller, 1991; Martini and Ryscavage, 1991). More changes would be reported in periods close to the current interview date than at the start of the recall period. While this may be related to the complexity of the information the respondent had to remember, it is probably also associated with other characteristics that affect the respondent's memory.

In studying the employment status seam effects in the PSID, Hill (1987) found that African Americans and older respondents were more likely to give inconsistent reports. He did not find any significant effect for sex, education or income. Jäckle and Lynn (2007) supported the view that older people where more likely to have inconsistent reports using BHPS employment data. In contrast to Hill, Callegaro and Belli (2007), also using PSID data to study employment status seam effects, found the reverse effect for age – older people were more likely to have consistent reports. They also found that people with lower education levels, females, and those below the poverty level were more likely to give inconsistent data. Jäckle (2008a) also observed that respondents with post-primary school qualifications make fewer recall errors.

Interview characteristics

The continuity of the interviewer may help the respondent recall their situation correctly. This may occur because the interviewer remembers some aspect of the prior interview or the very presence of the same interviewer triggers the respondent's memory. Different interviewers may also have subtle differences in their interviewing technique, such as the extent of their probing (Vick and Weidman, 1989).

The quality of the respondent's answers is also dependent on their understanding of the questions and their willingness to retrieve and interpret the required information from their memory. The respondents may simplify the task requested of them or avoid memory recall entirely by extrapolating their current situation to the past (either exactly or with some adjustments), thus resulting in constant wave response (Tourangeau *et al.*, 2000). Jäckle (2008a) found weak support for the hypothesis that co-operative respondents (as scored by the interviewer) had smaller seam effects. Respondents may also learn from previous interviews that certain responses lead to a series of questions on a specific topic and may opt for the response that avoids these (Burkhead and Coder, 1985).

Survey process characteristics

Seam effects can also result from changes or errors in the survey process. The question format, wording and order is important for consistent interpretation over time (Burkhead and Coder, 1985; Jäckle, 2008a; Callegaro and Belli, 2007). Interviewers may misunderstand or mis-record the information provided by the respondent (Burkhead and Coder, 1985; Lemaître, 1992). Data entry and coder inconsistencies (for example, in coding occupation) could result in discontinuities at the seam (Burkhead and Coder, 1985; Martini and Ryscavage, 1991; Halpin, 1998). Further, methods used to resolve discrepancies in the reported data from wave to wave can also contribute to the seam effect (Cotton and Giles, 1998).

3. The HILDA Calendar

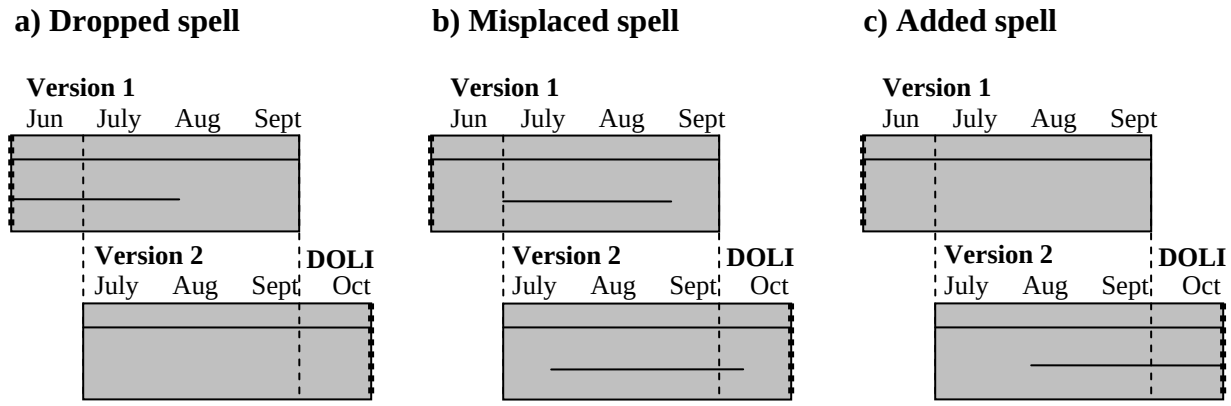
We now examine the recall errors made by respondents in the labour market activity calendar of the HILDA Survey.[1]

The HILDA Survey is an Australia-wide household panel survey, interviewing around 7000 households and 13,000 individuals each year (Watson and Wooden, 2004). The survey began in 2001 and each wave respondents are asked to recall the various jobs they have had, time spent in unemployment and time spent not in the labour force. In addition, spells of full-time and part-time education are also collected. Start and finish times of each spell are recorded by whether they are at the start, middle or end of each month. The calendar covers all months from 1st July of the preceding year to the date of the current interview (covering between 14 and 18 months).[2] All job spells should be included on the calendar (not just the main spell). Periods of full-time or part-time education can overlap with other types of spells. Where two or more jobs are recorded as having occurred at the very beginning of the calendar, their start dates are also collected to assist matching across waves.

As the calendar is administered every wave, an overlap of 2 to 6 months results, depending on when the respondent was interviewed. This overlapping period is used to identify spells that have been dropped, misplaced, or added as shown in Figure 1. In the first version of events, the calendar extends to the left (shown with the heavy dashed line), whereas in the second version of events, the calendar extends to the right. Two vertical lines have been drawn through both calendars in a light dashed line to show the 1st of July and the date of last interview (DOLI). In each of these examples, we have 3 months of overlap and one spell that matches perfectly. In the first example, the spell finishing at the beginning of

[1] The dataset used for the analysis in this paper is the In-Confidence Release 7 HILDA data.
[2] Interviews are conducted from mid August to March, with 96 to 98 per cent of the interviews completed by December.

Figure 1: Spell Recall Errors



August has not been reported in the second interview (i.e., it has been dropped). In the second example, the spell starting at the beginning of July has been moved (misplaced) to the end of July while adding an extra one third of a month to the duration. In the third example, a new spell has been inserted in the second interview in mid August (i.e., it has been added).

How these two versions of events will be resolved will depend on whether we match the spells at July 1st, the date of last interview or whether we attempt to reconcile the spells that have occurred in the overlapping period.

The grid used to collect this spell information in the HILDA Survey during the individual interview is provided in Figure 2 (the example is from wave 7). The wave 1 calendar was slightly different and aimed to collect whether jobs were full- or part-time or a mixture but this was not particularly successful and was dropped in later waves.

Table 1 provides a brief comparison of the key aspects of the labour market activity calendars collected in BHPS, German Socio-Economic Panel (GSOEP), PSID and HILDA. In the design of the HILDA calendar, key elements were picked from each of the other surveys. Like the BHPS and PSID, the HILDA calendar has an overlapping seam – it's median duration is slightly longer than the BHPS but not as long as the PSID. The grid of activities can be completed by moving forwards in time or backwards in time, depending on the respondent's preferences as is done in the GSOEP (in contrast to this, the BHPS and PSID work through a series of scripted questions). Also reflecting the GSOEP, only minimal information is collected about each activity due to time restrictions so we do not have access to information such as industry, occupation, wages, etc. to help match the job spells. Like the PSID, we

Figure 2: Labour activity calendar used in the HILDA Survey

Calendar

E9 I am now going to go over your work and study activities again so I can record these on a calendar.

NOTES: - *Record data for the period up to, and including, time of interview.*
- *The 3 boxes for each month represent the start, middle and end of the month.*
- *Holidays should not be treated as a break in the usual activity.*

a. Since July 2006, have you been enrolled in school or any course of study? IF YES: Was that full-time or part-time?

Yes ..... 1 - FILL IN CALENDAR FOR ALL TIME PERIODS ENROLLED
No 2

- *F/t or p/t study is determined by enrolment status (not hours).*
- *Only record courses or study that lead to a qualification.*

b. Since July 2006, how many different jobs (employers) have you had? No. of jobs

And what period did you work in [each / that] job? FILL IN TIME PERIOD FOR EACH JOB.
- *Use a new line for each new job / employer (no. of lines used must match no. of jobs).*
- *For periods of paid leave (e.g., long service, paid maternity leave) record as employed.*
- *If more than 1 job at start of calendar, record initial start dates for each job.*

At any time since July 2006, have you been:

c. … not employed BUT looking for work?

Yes..... 1 - FILL IN CALENDAR AT CODE 15
No 2

d. … not employed and not looking for work?

Yes..... 1 - FILL IN CALENDAR AT CODE 16
No 2

		2006						2007											
		July	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
	Enrolled in school / educational course																		
01	On FT basis																		
02	On PT basis																		
	Record start date (month and year) if more than 1 job at start of calendar.																		
03	Mth: ____ Yr: ____																		
04	Mth: ____ Yr: ____																		
05	Mth: ____ Yr: ____																		
06	Mth: ____ Yr: ____																		
07	Mth: ____ Yr: ____																		
08	Mth: ____ Yr: ____																		
09	Mth: ____ Yr: ____																		
10	Mth: ____ Yr: ____																		
11	Mth: ____ Yr: ____																		
12	Mth: ____ Yr: ____																		
13	Mth: ____ Yr: ____																		
14	Mth: ____ Yr: ____																		
15	Not employed but looking for work																		
16	Neither employed nor looking for work																		

NOTE: THERE NEEDS TO BE A MARK AGAINST EVERY TIME PERIOD IN THE CALENDAR PRIOR TO THE DATE OF INTERVIEW

Table 1: Labour market activity calendars in BHPS, GSOEP, PSID, and HILDA

	BHPS	GSOEP	PSID	HILDA
Period covered in calendar	1 Sept of proceeding year to date of interview	12 months in proceeding year	1 Jan two years prior to date of interview	1 July of proceeding year to date of interview
Fieldwork period	Sept-Dec (most waves)	Jan-Nov (96-99% generally completed by July)	March-Nov	Aug-March (96-98% completed by Dec)
Overlap in spells recalled	1 to 9 months (median 2 months)	None	3 to 11 months	2 to 6 months (median 2.7 months)
Activity recorded at each time point	Main only	Multiple categories. Some allowance for multiple jobs.	Main and others	Main and others
Level of detail in start and end dates	Exact date, but most analysis collapses this to month (Maré, 2006)	Month	Month and third of month	Month and third of month
Type of instrument	List of questions going back in time from current spell	Grid of activities which can be completed forwards or backwards	Event history calendar (using other significant events to help recall) See Belli *et al.* (2001).	Grid of activities which can be completed forwards or backwards.
Details collected about the spells	Industry, occupation, wage, etc.	None	Industry, occupation, type and size of company, wage, hours, etc	None
Questionnaire	Pages 83-89 of the 2007 Individual Questionnaire[3]	Page 23 of the 2008 Individual Questionnaire[4]	Pages 7-17 of the 2007 Questionnaire[5]	Page 17 of the 2007 Continuing Person Questionnaire and Page 21 of the 2007 New Person Questionnaire[6]

aim to capture all spells, not just the main activity spells that are recorded by the BHPS (so this eliminate some of the matching problems faced by the BHPS due to changes in what was considered the 'main' activity). Like the PSID, the HILDA Survey collects the start and end dates for the spells in thirds of a month (beginning, middle and end) as this tends to be closer to how the respondent remembers the spell (rather than say week 1, 2, etc. or trying to collect exact start and end dates).

In the HILDA calendar, we do not have any missing dates reported. The respondents and interviewers resolve any missingness in dates by providing their 'best guess' of what happened. The design of the

[3] http://www.iser.essex.ac.uk/survey/bhps/documentation/pdf_versions/questionnaires/bhpsw17q.pdf
[4] http://www.diw.de/documents/dokumentenarchiv/17/85645/personen_2008_en.pdf
[5] ftp://ftp.isr.umich.edu/pub/src/psid/questionnaires/q2007.pdf.
[6] http://www.melbourneinstitute.com/hilda/qaires/ContinuingPersonQuestionnaireW7.pdf and http://www.melbourneinstitute.com/hilda/qaires/NewPersonQuestionnaireW7.pdf.

calendar may, therefore, cause some inconsistencies between the two reports. Further, in the first seven waves of the HILDA Survey, only two respondents on one occasion each have refused to complete the calendar.

The following table provides some more information on the overlap in the HILDA calendar data for people interviewed in all seven waves. With two exceptions, each person contributes six overlapping seams (two people contributes five overlapping seams). The length of the overlap ranges from 1 to 6 months, with the average being 2.9 months.

Table 2: Overlap in the HILDA activity calendar

Length	Frequency	Percent
1 month (3-4 thirds)	101	0.2
2 months (5-7 thirds)	16,642	33.0
3 months (8-10 thirds)	24,352	48.3
4 months (11-13 thirds)	6,211	12.3
5 months (14-16 thirds)	2,497	5.0
6 months (17-18 thirds)	649	1.3
Total	50,452	
Mean	2.9 months	
Median	2.7 months	

4. Matching Methods

To investigate the value of the overlapping seam, three matching methods are evaluated using the first 7 waves of the HILDA data:

i) Match the jobs at 1st July (allowing for the maximum recall errors to be seen);

ii) Match the jobs at the date of last interview (giving precedence to the information collected closest to the period being recalled);

iii) Reconcile spell reports between the two versions of events reported approximately one year apart.

In the first two matching methods, the alternative version of spells reported for the same period is ignored. Where multiple spells from one wave could be matched to one spell in another (as occurs with job spells), a match was randomly chosen. Spells are matched within each type of spell (being full-time education, part-time education, job, unemployment, and not in the labour force).

The third matching method seeks to use the two reports to produce a reconciled view of the overlapping period. The first version of events (recalled closest to the period when they occurred) is taken in precedence over the second. However, spells recorded in the second version of events are matched with those in the first to identify (and remove) inconsistent reports that result in a job spell being misplaced. The second report is assumed to be incorrect (as it occurs further away from the period when the spell occurred). This method cannot resolve any situation where a spell in the first version is failed to be recalled in the second version of events or visa versa. Spells are matched within each type of spell. A match score was created to help identify which spells should be matched to which. The match score is the sum of:

i) 1 if the start of the spells match exactly within the overlapping seam;

ii) 1 if the start of the spells match within one month;

iii) 1 if the end of the spells match exactly within the overlapping seam;

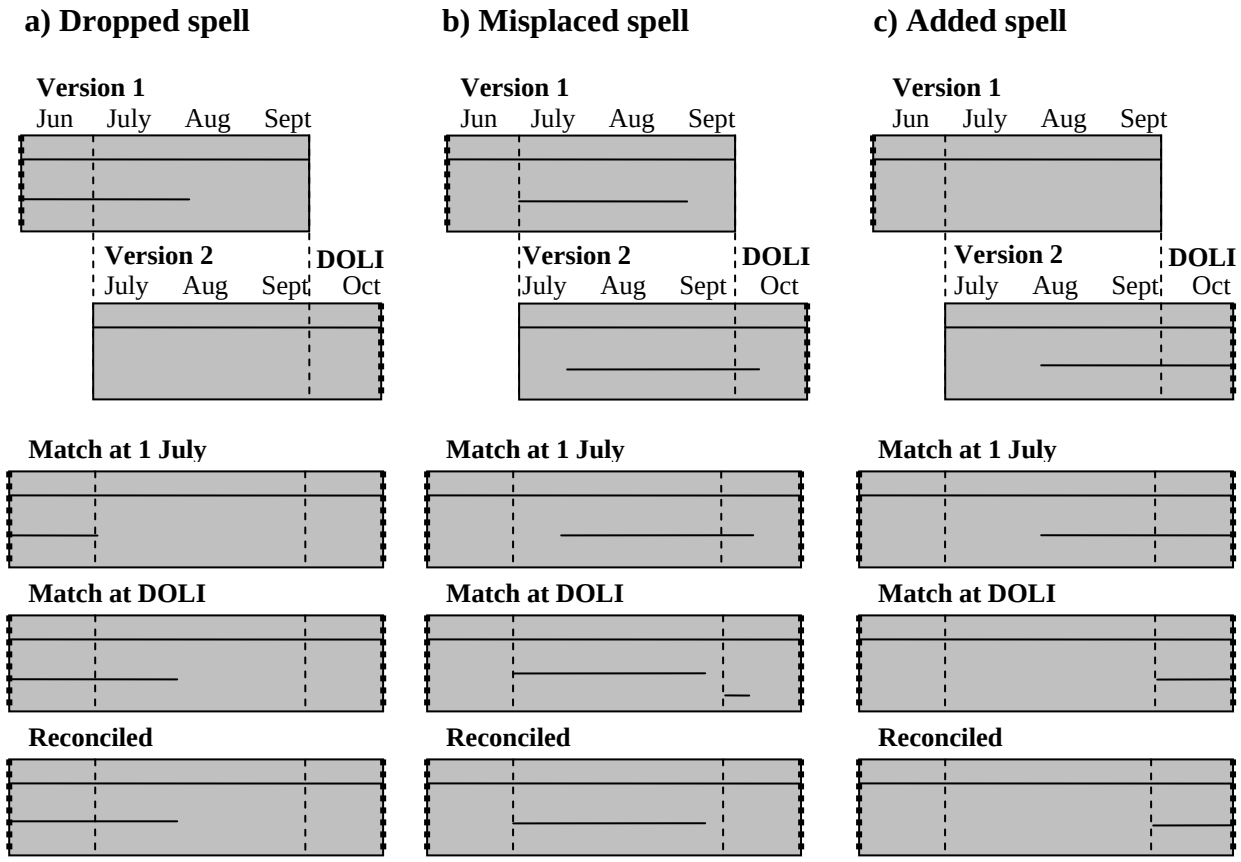
iv) 1 if the end of the spells match within one month; and

v) 3 times the ratio of the length of the spells (with the longest in the denominator).

The maximum score for a match is therefore 7. Spells with the highest match score are matched first. The remaining spells are then matched in the same way until no more spells can be matched. Spells with a match score of less than 3 are not matched (for example a spell of one third of a month in July could not be matched to one in the second version of events to one occurring for all of September).

Figure 3 shows how each of the examples provided in Figure 1 of dropped, misplaced or added spells would be resolved under these three matching methods.

Figure 3: Matching implications for spell recall errors



5. Results

Size of the seam effect

The number of job transitions at the seam for spells matched at the start of July is around eight times those off-seam. For the other two methods, the seam is at the date of interview of each respondent, so spreads from mid August to December. An elevation in the number of job transitions is observed over this period. Figure 4 shows the (unweighted) number of job starts and ends for those interviewed in each of the first seven waves of the HILDA Survey. The black line shows the number of job starts when the spells are matched at the start of July (method 1 [LFY]). The red line shows the results for when the spells are matched at the date of last interview (method 2 [DOLI]) and the barely visible grey line is when the spells are reconciled (this method 3 [OLAP] behaves very much like method 2).

Several observations can be made about these graphs:

- Respondents tend to report fewer job starts and job ends at the beginning of the calendar (this can be seen in the results for method 1 (black line) between July and October each year. This

suggests either that there is decay in the memory over time or that respondents are backwards telescoping events to the boundary of the reference period (a similar effect was found by Kalton and Miller (1991) in an analysis of data from the US Survey of Income and Program Participation). When spells for this overlapping period are taken from the respondent's first report of the events, this difference is eliminated (this can be seen in July and early August in the DOLI and OLAP figures before the effect of the seam at the date of last interview occurs in figures for mid August to December).

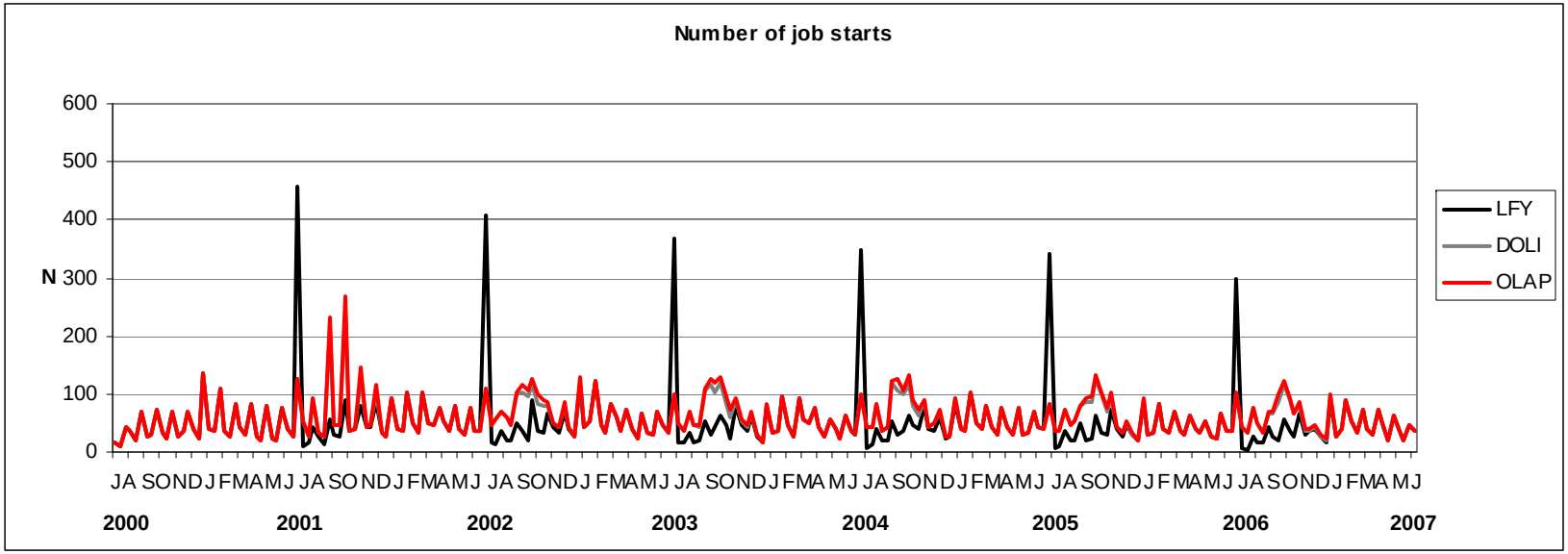
- Respondents tend to report jobs starting at the beginning of a month and ending at the end of the month, resulting in a spike in the graph each month. While there might be a real effect for jobs to start and finish in this way, it is probably accentuated by recall error. Note that this tendency seems to be become slightly less apparent in the later waves.

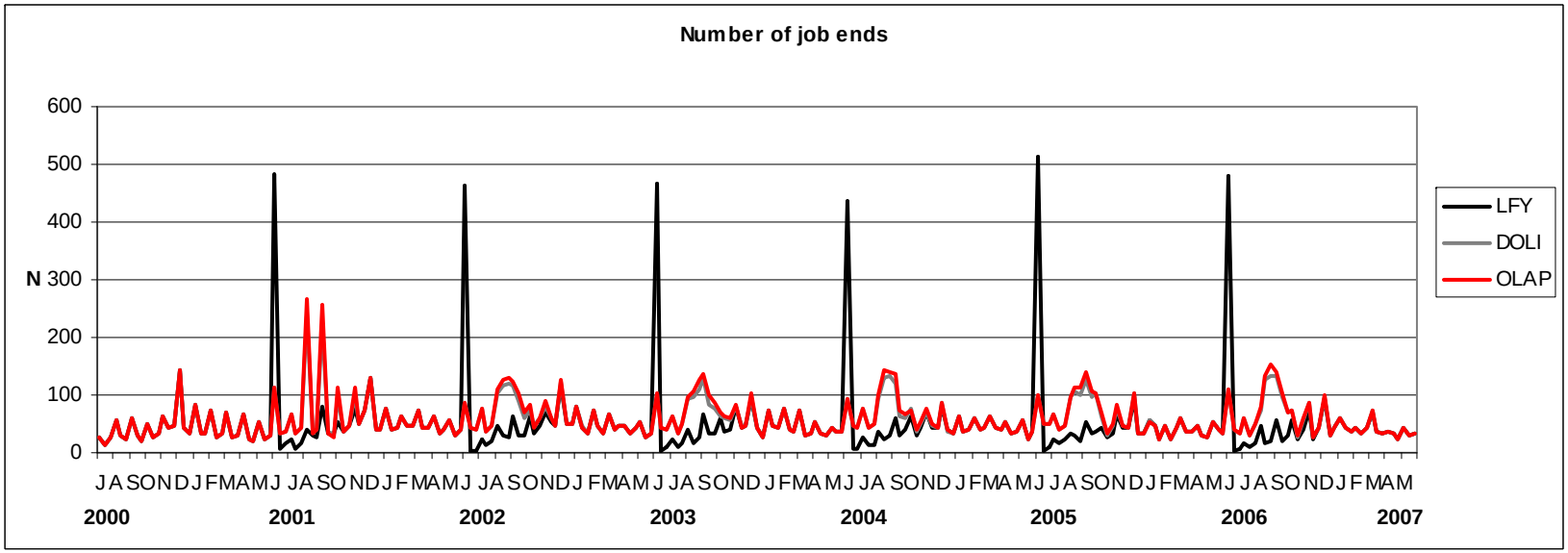
- More job starts seem to occur at the start of July and January than other months and end at the end of June and December. Again, this might be a real effect, particularly for people working on contracts. (The June/July effect can be seen in the results for DOLI and OLAP methods but this effect is obscured in the LFY method by the seam effect.)

- There is also more heaping of job starts and ends for the DOLI and OLAP methods at the end of wave 1 (around August to November 2001) due to instructions given to the interviewers in wave 1 to 'record activities up until the end of the month prior to the date of the interview'.

- The differences between DOLI and OLAP methods are very minor, though we do find that the OLAP method tends to have marginally higher transitions during the main fieldwork months (September and October).

Figure 4: Number of job starts and job ends, July 2000 to June 2007





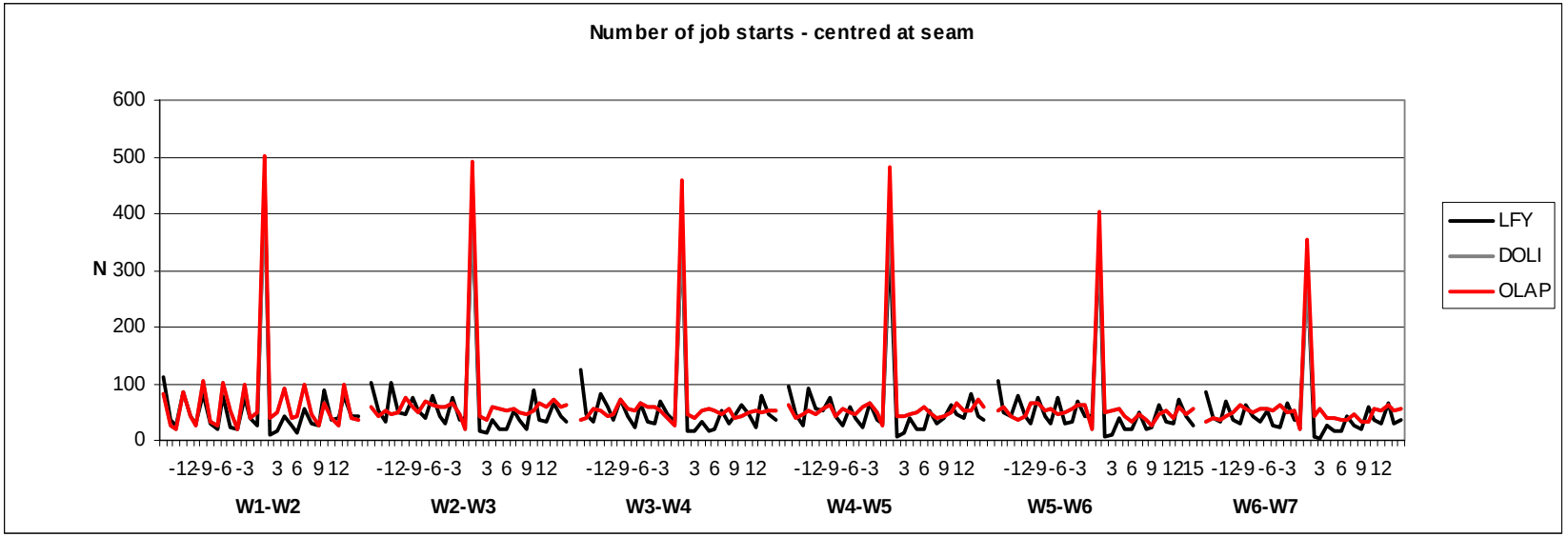
To help identify whether DOLI and OLAP methods actually reduce the effect of the seam (or just disguise it by spreading it across the interview period), Figure 5 centres the count of job starts and ends at the seam. For clarity, the counts of job starts and ends for each method at the seam are also reported in Table 3. Several observations from this figure and table are:

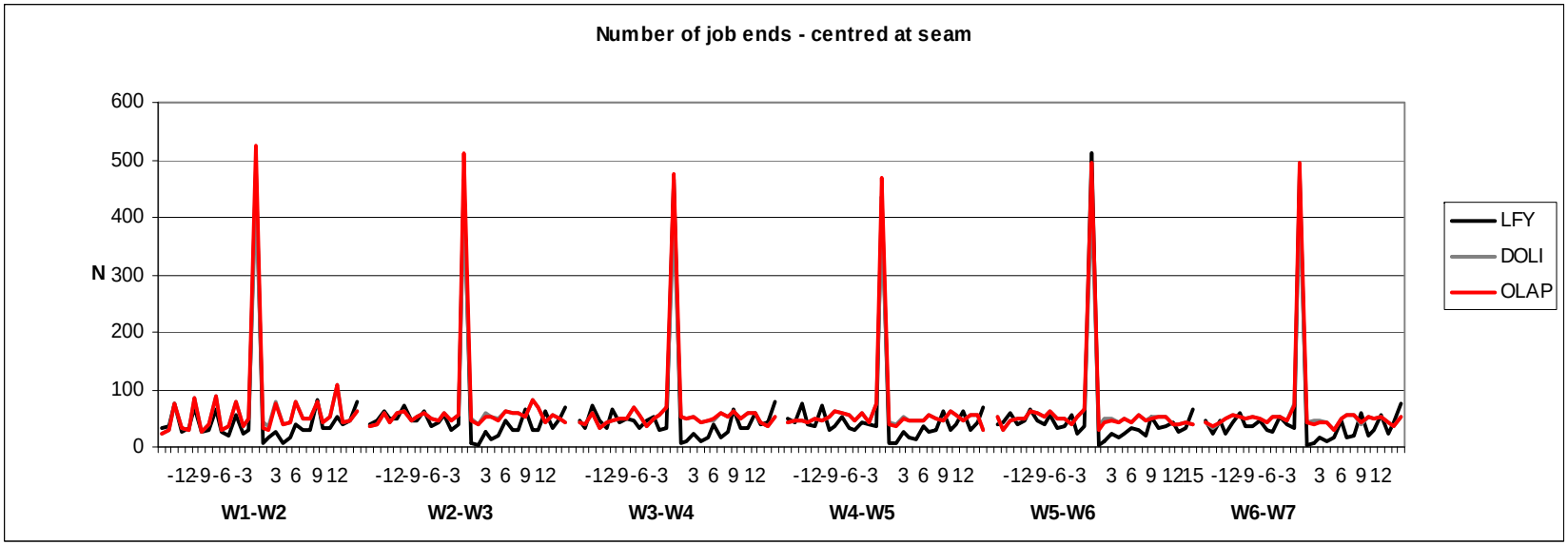
- DOLI method has slightly more job starts at the seam than LFY method simply because there are more jobs reported later in the calendar than earlier.
- The OLAP method has more job starts and ends at the seam than the DOLI method. When comparing the spells in the two versions of the overlapping seam, it is apparent that some of the spells matched in the DOLI method are actually not the same so the spells are not connected, this results in more job ends and job starts at the seam.
- The number of job starts at the seam appears to decline over time, but there is not a corresponding decline in the number of job ends at the seam. It is not apparent what might be causing this.
- It is clear that the DOLI and OLAP methods are better than the LFY method as these methods avoid the drop in the number of spells that start or end at the beginning of the reference period. The OLAP method is better than the DOLI method as it helps us identify spells that should not be matched in the overlapping period.

Table 3: Number of job starts and ends at the seam, by wave and method

	start			*end*		
	Method 1: LFY	*Method 2: DOLI*	*Method 3: OLAP*	*Method 1: LFY*	*Method 2: DOLI*	*Method 3: OLAP*
W1-W2 seam	459	426	502	485	432	526
W2-W3 seam	407	410	493	465	416	512
W3-W4 seam	368	372	458	467	387	474
W4-W5 seam	347	405	482	437	389	469
W5-W6 seam	342	350	404	512	423	495
W6-W7 seam	298	293	353	480	421	495
Total	2,221	2,256	2,692	2,846	2,468	2,971

Figure 5: Number of job starts and job ends, centred at seam between each wave





Types of recall errors

Respondents can drop, misplace or add spells within the seam, but the effect this has on the seam in the DOLI or OLAP methods depends on whether the spells cross the right edge of the overlapping period or not. Table 4 summarises the spells reported in the overlapping period for each type of spell. Take for example the job spells. We find that 83 per cent of the spells reported in the first version of the overlapping period match exactly, 12 per cent are dropped and 5 per cent are misplaced. Of all the job spells reported in the second version of the overlapping period, 8 per cent are added. The number of spells that are added is about two thirds of the number of spells dropped (as would be expected if respondents are more likely to forget spells that occurred longer ago). Almost all (99 per cent) of the job spells that are matched exactly cover the entire overlapping period. Nearly 70 per cent of the job spells that are dropped end at the right hand edge of the overlapping period and thus contribute to the seam effect. More than 90 per cent of the spells that are added cross the right hand edge of the overlapping period and also contribute to the seam effect.

The proportion of spells that match exactly in the overlapping period varies greatly by the type of spell. For job spells, the proportion matching exactly (as mentioned above) was 83 per cent. Slightly more spells not in the labour force were matched correctly (85 per cent). Two thirds of the full-time education spells, 40 per cent of the part time education spells and only 24 per cent of the unemployment spells matched exactly.

Sometimes in the matching process, it is completely impossible to match the spell reported in the first version of the overlapping seam as there is no corresponding spell of that type in the second version. This occurred for approximately 290 job spells, 330 not in the labour force spells, 280 unemployment spells, 190 full-time education spells, and 340 part-time education spells. Table 5 shows how these unmatched spells are redefined. Job spells are most likely to be redefined to time not in the labour force and visa versa. Unemployment spells are also more likely to be redefined to not in the labour force rather than job spells. For education spells, both full-time and part time spells are almost always redefined as not studying (and almost never swaps between full- or part-time education).

Table 4: Recall errors in the overlapping seam

Type of spell	*Placement*	*Exact match*	*Dropped*	*Misplaced*	*Added*
Job	Entire seam	31,120	2,334	252	2,228
	Not at right edge of seam	179	1,414	453	219
	Only at right edge of seam	249	772	1,148	428
	Total	31,548	4,520	1,853	2,875
Unemployed	Entire seam	532	896	31	598
	Not at right edge of seam	26	110	93	118
	Only at right edge of seam	19	563	166	134
	Total	577	1,569	290	850
Not in labour force	Entire seam	14,922	1,291	89	1,341
	Not at right edge of seam	32	254	148	182
	Only at right edge of seam	26	511	274	150
	Total	14,980	2,056	511	1,673
Full-time education	Entire seam	1,815	596	28	388
	Not at right edge of seam	9	85	29	18
	Only at right edge of seam	3	91	51	26
	Total	1,827	772	108	432
Part-time education	Entire seam	1,369	1,446	49	787
	Not at right edge of seam	6	86	22	56
	Only at right edge of seam	18	347	131	133
	Total	1,393	1,879	202	976

Table 5: Reclassification of spells, by method

		Per cent
Labour force	job -> not in labour force	76.3
	job -> unemployed	23.7
	not in labour force -> job	63.9
	not in labour force -> unemployed	36.1
	unemployed -> job	33.9
	unemployed -> not in labour force	66.1
Education	full-time -> part-time	0.5
	full-time -> none	99.5
	part-time -> full-time	0.3
	part-time -> none	99.7

Characteristics associated with recall errors

The overlap period is used to identify:

i) spells reported in the first version of events and not in the second (dropped spells);

ii) misplaced spells in the second version of events compared to the first; and

iii) spells reported in the second version of events and not in the first (added spells).

The spell-level dataset is pooled across the first seven waves of the HILDA Survey and is restricted to individuals who were interviewed all seven waves. This allows us to consider how well the spells match across six overlapping seams. Only those spells occurring in the overlapping seam are included in the analysis. Three separate logistic models were used to estimate the likelihood that a spell would be dropped, misplaced or added. The pooled dataset of the first version of calendar events contains 64,085 spells on 8409 individuals across six seams. This dataset was used to estimate the likelihood of dropping or misplacing spells. A second pooled dataset of the second version of calendar events contains 59,981 spells on 8409 individuals across six seams and this was used to estimate the likelihood of adding spells. To allow for the repeated measures on the same individuals, the estimated standard errors assume the spell outcomes are correlated across observations on the same individual but are independent across individuals. The models were estimated separately for each of the different spell types.

Using the structure presented in section 2 above, the probability that a spell was dropped, misplaced or added was assumed to be a function of: the spell characteristics, the complexity of the recall, characteristics of the respondent; the interview characteristics and survey process characteristics.

The *spell characteristics* are represented by the spell length (counted in thirds of a month) and whether the spell was the same as an activity recorded for the date of interview relating to the second version of events.

The *complexity of the recall* was estimated by the number of spells recorded in the calendar between the last date of interview and the current interview, and the time between the two interviews. The length of the overlap increases the opportunity for spells to disappear or appear (Paull, 2002), so this variable has been included in the model as a control.

The *respondent characteristics* included in the model were sex, age, education level, financial year personal income, country of birth and whether an indigenous Australian. The age groups used in the analysis of the education spells are 15-24 and 25 and over, whereas the age groups used for the other spells are 15-24, 25-54, 55 and over.

The *interview characteristics* incorporated into the model were whether the interviewer rated the respondent's understanding of the questions as 'excellent' or 'good', whether the interview was conducted by telephone and whether the same interview conducted both interviews. A measure of the interviewer's experience was included via a count of the number of calendars the interviewer had completed that wave (with the expectation that the more calendars completed, the better they would be at obtaining sufficient detail from the respondent).

The *survey process characteristic* included in the model was whether the seam related to that between waves 1 and 2. This was included as there was a change in the calendar design between waves 1 and 2. In wave 1, each job was meant to be separated into whether it was full-time or part-time, however interviewers found this difficult to remember to record correctly. Indeed, in a few calendars it appears that some interviewers may have sometimes collapsed all jobs into just two job spells – one for part-time and one for full-time. A simplified version of the calendar was used from wave 2 onwards.

The results of the model estimation for job spells, unemployed spells and the spells not in the labour force are presented in Table 6. The following observations are made about these three models:

i) The single most important factor in making a spell recall error is whether the same type of activity was current at the date of the interview for the second recall period. If the same type of activity was also a current activity, the likelihood of dropping or adding a spell was significantly reduced. For example, the mean predicted probability of dropping a job spell is 7.5 per cent for those with a job at the date of interview but is 31.1 per cent for those who do not. For job spells, it also reduced the likelihood of misplacing the spell, but for spells of unemployment or not in the labour force it *increased* the chances of misplacing the spell.

ii) Consistent with previous research, shorter spells are more likely to be dropped, added or misplaced. The exception noticed in the HILDA data is that the length of an

unemployment spell does not affect the probability of it being added to the activity calendar.

iii) In the main, the complexity of the recall task (as measured by the number of spells to be recalled) increases the probability of making errors. Contrary to expectations the HILDA data shows that with greater complexity in the recall, unemployment and not in the labour force spells are *less* likely to be dropped.

iv) There is some support for the length of the recall period and the overlapping seam to increase the chances of a respondent making a recall mistake.

v) The models suggest that respondents tend to make mistakes with spells that are more unusual for their time of life. Older respondents were more likely to make mistakes with job spells but are less likely to make mistakes with spells not in the labour force. Women are more at risk of making mistakes with unemployment spells than men – perhaps these spells are more fuzzy for women as they may reclassify their unemployment time as time not in the labour force. Young respondents are more likely to make mistakes with spells not in the labour force.

vi) Education, country of birth and indigenous status often had no effect on the likelihood of recall errors.

vii) With higher individual incomes, the likelihood of dropping or adding job spells is increased, but the changes of making such mistakes with spells not in the labour force is reduced.

viii) Few of the interview characteristics had an effect on the likelihood of making a recall error. Respondents who appeared to understand the interview questions were less likely to add job spells or drop spells not in the labour force. The estimates from the models suggest that people interviewed by telephone could be more likely to make mistakes with the calendar (while not significant at the 5 per cent level, the mean predicted probability of dropping a job spell increased by 1.3 percentage points when the interview was conducted by telephone).

Table 6: Characteristics associated with recall errors in labour force spells

	Job spells			*Unemployment spells*			*Not in labour force spells*		
Variable	*Drop*	*Misplace*	*Add*	*Drop*	*Misplace*	*Add*	*Drop*	*Misplace*	*Add*
Spell characteristics									
Spell length	-0.035***	-0.106***	-0.013***	-0.024***	-0.067***	0.004	-0.048***	-0.103***	-0.025***
Same type of spell at date of interview	-1.840***	-0.521***	-1.279***	-1.527***	0.520***	-0.632***	-4.360***	0.295*	-2.952***
Complexity of recall									
Number of spells btw ivws	0.057***	0.085***	0.308***	-0.158***	0.105**	0.089*	-0.352***	0.181***	-0.040
Time between last and current ivw date	0.028***	-0.032**	0.010	-0.012	-0.018	0.013	0.020	-0.019	0.038*
Length of overlap	0.072***	0.138***	0.057***	0.005	0.050	0.074**	0.043	0.162***	0.062***
Respondent characteristics									
Sex and age (base = younger males aged 15-24)									
Prime males (aged 25-54)	-0.168*	0.282**	-0.081	-0.052	-0.085	-0.095	0.226	-0.057	-0.357
Older males (aged 55 or older)	0.210*	0.006	0.393***	0.419*	-0.620	0.088	-0.140	-0.023	-1.091***
Younger females	0.065	0.149	-0.171	0.392	-0.299	0.281	-0.090	0.515	-0.221
Prime females	-0.069	0.091	0.103	0.316	-0.072	0.298	-0.138	0.507*	-0.844***
Older females	0.119	0.303*	0.408***	0.933***	-0.663	0.778***	-0.312	0.093	-1.381***
Education level (base=year 11 or below)									
Year 12	0.026	0.028	0.005	0.257	-0.181	0.227	0.084	0.194	0.041
Certificate	0.117*	0.046	-0.087	0.099	-0.359	0.056	0.132	0.475***	0.184*
Diploma	0.095	-0.115	-0.030	0.268	-0.557	0.264	-0.018	0.064	-0.082
Graduate	0.203***	0.004	-0.017	-0.051	-0.206	0.158	0.099	0.082	0.227**
Financial year income ($/10^5$)	-0.527***	0.175	-0.568***	-1.022	1.210	0.182	1.506***	-0.045	1.277***
Financial year income squared ($/10^{10}$)	0.050**	-0.051	0.048***	0.329	-0.429	0.034	-0.245***	-0.227	-0.273*
Country of birth (base = Australia)									
Main English-speaking country	-0.127	-0.023	-0.090	0.048	0.221	0.144	0.040	0.100	0.010
Other overseas country	0.116*	-0.159	0.121	0.037	0.103	-0.101	-0.053	0.298*	0.004
Indigenous Australian	0.170	-0.023	0.215	0.286	-1.389*	0.021	0.386	-0.249	-0.052
Interview characteristics									
Understanding of questions	-0.175	0.072	-0.403***	0.065	-0.171	0.008	-0.442***	0.247	0.051
Telephone interview	0.168*	-0.022	-0.004	0.120	0.229	-0.100	0.205	-0.319	0.380*
Continuity of interviewer	-0.094**	-0.116*	-0.043	-0.086	0.004	-0.126	-0.072	-0.009	-0.161**
Experience of interviewer (calendars completed)	0.000	0.001	-0.001	0.001	0.000	-0.001	0.000	0.000	-0.001
Survey process characteristics									
Seam between wave 1 and 2	-0.166**	-0.331***	0.107	-0.224	0.028	0.283*	0.035	0.182	0.012
Constant term	-0.880*	-0.354	-2.043***	1.718	-1.013	-1.334	1.619	-2.577*	-0.509
N spells	36,383	36,383	34,529	1,771	1,771	1,412	16,843	16,843	16,744

Note: * significant at 10 per cent, ** significant at 5 per cent, *** significant at 1 per cent.

ix) There was some evidence that interviewer continuity helped reduce the likelihood of making recall errors, but the effect was relatively small (for example, the mean predicted probability of dropping a job spell decreased by 0.7 percentage points when the interviewer was the same). The experience of the interviewer was not significant, regardless of whether this was estimated as the number of calendars completed each wave or over all waves.

x) We did not find support for the hypothesis that job spells were more likely to be added in the overlapping period between waves 1 and 2. Unexpectedly, we found that job spells were *less* likely to be dropped or misplaced in this overlapping period. It is not clear why this might be the case.

The results of the model estimation for full-time and part-time education spells are broadly similar (shown in Table 7). The length of the spell, whether the same type of spell occurs at the date of interview, the number of spells in the recall period and the length of the overlapping period are important to whether a recall error is made. Nevertheless, there are some important differences:

i) The differences by age are much more apparent for the education spells, with older respondents being much more likely to drop or add full-time education spells. Young females were more likely to drop part-time education spells than young males.

ii) More educated respondents are less likely to drop or add education spells. They are also likely misplace part-time education spells.

iii) Respondents born overseas in a non-English speaking country are more likely to misplace full-time education spells and indigenous Australians are more likely to drop or add part-time education spells.

iv) While the continuity of the interviewer was not significant, the more interviews they had completed the less likely they were to have respondents who drop or add full-time education spells. Part-time education spells are more likely to be added as the interviewer gains more experience with the calendar. Both of these findings suggest that interviews become better at probing for spells over time.

Table 7: Characteristics associated with recall errors in education spells

	Full-time education spells			*Part-time education spells*		
Variable	*Drop*	*Misplace*	*Add*	*Drop*	*Misplace*	*Add*
Spell characteristics						
Spell length	-0.017***	-0.088***	0.024***	-0.017***	-0.118***	-0.020***
Same type of spell at date of interview	-3.417***	-0.428	-1.885***	-2.385***	0.754***	0.439***
Complexity of recall						
Number of spells btw ivws	-0.373***	0.187**	-0.219***	-0.496***	0.311***	-0.041
Time between last and current ivw date	0.018	0.029	0.061*	-0.004	-0.062	-0.009
Length of overlap	0.059*	0.254***	0.117***	0.058**	0.136**	0.028
Respondent characteristics						
Sex and age (base = younger males aged 15-24)						
Prime and older males (aged 25 or older)	0.646***	-0.263	0.494**	0.244	-0.249	-0.241
Younger females	0.030	0.009	-0.172	0.567**	0.184	0.250
Prime and older females	0.517***	-0.311	0.623***	0.025	-0.263	-0.607***
Education level (base=year 11 or below)						
Year 12	-0.098	-0.024	-0.156	-0.116	0.312	-0.002
Certificate	-0.283	0.587	0.494*	-0.735***	-0.047	0.018
Diploma	-0.872***	0.568	-0.835***	-1.058***	0.774**	-0.065
Graduate	-1.251***	0.500	-0.730***	-0.893***	0.333	-0.275*
Financial year income ($/10^5$)	3.383***	0.772	5.196***	-0.649***	1.337**	-1.267***
Financial year income squared ($/10^{10}$)	-1.771***	-2.682	-2.479***	0.191***	-0.305	0.484***
Country of birth (base = Australia)						
Main English-speaking country	0.299	0.238	-0.026	0.316*	-0.167	0.189
Other overseas country	-0.025	0.696**	0.267	-0.019	-0.057	0.145
Indigenous Australian	0.365	0.221	0.278	1.383***	-0.528	0.756*
Interview characteristics						
Understanding of questions	-0.095	0.274	-0.470	-0.334	0.834	-0.300
Telephone interview	0.329	-0.770	0.325	0.041	-0.323	0.295
Continuity of interviewer	0.107	-0.132	0.063	-0.080	0.002	-0.134
Experience of interviewer (calendars completed)	-0.003**	-0.003	-0.003**	0.001	0.000	0.002*
Survey process characteristics						
Seam between wave 1 and 2	0.111	-0.761*	-0.227	-0.138	-0.105	0.214
Constant term	0.447	-4.615*	-4.195***	2.975***	-1.982	1.013
N spells	2,514	2,514	2,290	3,035	3,035	2,430

Notes: * significant at 10 per cent, ** significant at 5 per cent, *** significant at 1 per cent.

The above estimated models do not let us determine whether respondents make the same mistakes over time. To test this, the pooled dataset was restricted to the overlapping seams for waves 2 and 3 through to waves 6 and 7 and an additional variable indicating whether the respondent had made the same mistake (of dropping, misplacing or adding a spell) in the previous seam. We find that dropping a spell in the previous overlapping period increases the risk of dropping a spell in the following overlapping period. For example, the mean predicted probability for dropping a job spell increased by 1.9 percentage points when a spell of any type had been dropped in the previous overlapping seam. Similarly, if a spell is added in the previous overlapping period, the chances of adding a spell in the next overlapping period is higher for most types of spells. Errors in reporting full-time education spells

do not appear to be affected by the types of errors made in the past. Also, misplacement of spells in one overlapping period does not seem to have any bearing on whether spells are misplaced in the next overlapping period.

6. Discussion

The overlapping seam collected in the HILDA labour market activity calendar has not resolved the problem of the seam effect. By matching the spells at the date of last interview, each respondent effectively has their own seam and this spreads the seam effect out across the interviewing period. This analysis has shown that matching at the date of last interview is better than at the last financial year as this avoids the problems of low recall of spell starts and ends in the first few months of the calendar. The method which attempts to reconcile the spells across the seam identifies spells which are not actually the same and, even though it produces a higher number of dropped or added spells, the resultant spell file is more accurate.

The most common mistake made by respondents is to drop a previously reported spell, with 17 per cent of spells being dropped. Unemployment spells and part-time education spells are most at risk of being dropped. Respondents tend to add fewer spells than they drop, owing to poorer recall of events in the more distant past. The least common mistake is to misplace spells, with around 5 per cent of spells are misplaced.

The spells which are most subject to recall error are spells that are unlike those reported at the current date of interview, short spells, and those spells which are part of a complex history. Some limited support was found in this study for reduced recall error when the interview is conducted face-to-face, the interview is the same between waves, and the interviewer has greater experience in completing the calendars. Respondents who seemed to have a good understanding of the questions asked in the entire interview (as rated by the interviwer) were less likely to make some recall mistakes in the calendar. The effect that the respondent characteristics had on the likelihood of recall errors varied by the type of spell recalled. For example, respondents tend to make more mistakes with spells that are less congruent with their stage in life – older respondents were more likely to make mistakes with job spells and unemployment spells, but less likely to make mistakes with spells not in the labour force. We also found evidence that respondents tend to make the same recall mistakes over time in terms of dropping or adding spells, but not in misplacing spells. Users of the HILDA calendar data will need to be aware of these differences in recall error as these may have a bearing on their findings.

The production of a matched HILDA labour market activity spell file for release to users will require further work. Some recall errors may be resolved by comparing data from other parts of the individual interview (such as the tenure of the current job or time spent unemployed) with what has been reported in the calendar. Further, it will also be important to identify and correct any flow-on effects of accepting a misplaced spell match to ensure that spells do not overlap in inconsistent ways and that each period of the calendar has been covered by at least one spell. It would also be helpful to provide users with some guidance on how best to analyse the data in the presence of significant seam effects. As a starting point, Lynn *et al.* (2005) provides a summary of the current methods adopted by users to deal with seam effects.

It is also important to continue to investigate methods which will help minimize the seam effects in the data. Dependent interviewing is much more of a reality now with the switch to Computer Assisted Personal Interviewing in wave 9 of the HILDA Survey. It is, however, difficult to see how we could feed forward the information required for the HILDA calendar in a way that would make sense to the interviewer and the respondent. Nevertheless, other studies have found that dependent interviewing greatly reduces seam effects (Jäckle, 2008b; Moore *et al.*, 2009), so it is definitely worth considering.

References

Belli, R.F., Shay, W.L., and Stafford, F.P. (2001), 'Event History Calendars and Question List Surveys', *Public Opinion Quarterly*, Vol. 65, pp. 45-74.

Burkhead, D. and Coder, J. (1985), 'Gross Changes in Income Recipiency from the Survey of Income and Program Participation', *Proceedings of the American Statistical Association, Social Statistics Section*, Washington, DC, pp. 351-356.

Callegaro, M. (2008), 'Seam Effects in Longitudinal Surveys', *Journal of Official Statistics*, Vol. 24, pp. 387-409.

Callegaro, M. and Belli, R.F. (2007), 'Impact of Event History Calendar on Seam Effect in PSID: Lessons for SIPP', paper presented at Event History Calendar Methods Conference, Dec 5-6, Washington, DC.

Carroll, N. (2006), 'Explaining Unemployment Duration in Australia', *Economic Record*, Vol. 82, No. 258, pp. 298-314.

Cotton, C. and Giles, P. (1998), 'The Seam Effect in the Survey of Labour and Income Dynamics', SLID Research Paper Series, Cat. No. 98-18, Statistics Canada.

Dockery, A.M. (2004), 'Looking Inside the Unemployment Spell', *Australian Journal of Labour Economics*, Vol. 7, No. 2, pp. 175–198.

Halpin, B. (1998), 'Unified BHPS Work-Life Histories: Combining Multiple Sources into a User-Friendly Format', *Bulletin de Méthodologie Sociologique*, No. 60, pp. 34-79.

Hill, D.H. (1987), 'Response Errors Around the Seam: Analysis of Change in a Panel with Overlapping Reference Periods', *Proceedings of the American Statistical Association*, Section on Survey Research Methods, Washington, DC, pp. 210-215.

Jäckle, A. (2008a), 'The Causes of Seam Effects in Panel Surveys', Institute for Social and Economic Research Working Paper Series, No. 2008-14.

Jäckle, A. (2008b), 'Dependent Interviewing: Effects on Respondent Burden and Efficiency of Data Collection', *Journal of Official Statistics*, Vol. 24, No. 3, pp. 411-430.

Jäckle, A. and Lynn, P. (2007), 'Dependent Interviewing and Seam Effects in Work History Data', *Journal of Official Statistics*, Vol. 23, No. 4, pp. 529-551.

Kalton, G. and Miller, M.E. (1991), 'The Seam Effect with Social Security Income in the Survey of Income and Program Participation', *Journal of Official Statistics*, Vol. 7, No. 2, pp. 235-245.

Lemaître, G. (1992), 'Dealing with the Seam Problem for the Survey of Labour and Income Dynamics', in SLID Research Paper Series, Statistics Canada, Ottawa.

Lynn, P., Buck, N., Burton, J., Jäckle, A, Laurie, H. (2005), 'A Review of Methodological Research to Longitudinal Survey Design and Data Collection', ISER Working Paper Series, No. 2005-29, University of Essex, Colchester.

Maré, D. (2006), 'Constructing Consistent Work-life Histories: A guide for users of the British Household Panel Survey', ISER Working Paper Series, No. 2006-39, University of Essex, Colchester.

Martini, A. and Ryscavage, P. (1991), 'The Impact of Survey and Questionnaire Design on Longitudinal Labour Force Measures', paper at Annual Research Conference of the US Bureau of the Census, Arlington, VA.

Moore, J., Bates, N., Pascale, J. and Okon, A. (2009), 'Tackling Seam Bias Through Questionnaire Design', in Lynn, P. (ed), *Methodology of Longitudinal Surveys*, John Wiley and Sons, Chichester.

Paull, G. (2002), 'Biases in the Reporting of Labour Market Dynamics', The Institute for Fiscal Studies Working Paper Series, No. 02/10.

Rips, L.J., Conrad, F.G. and Fricker, S.S. (2003), 'Straightening the Seam Effect in Panel Surveys', *Public Opinion Quarterly*, Vol. 67, pp. 522-554.

Tourangeau, R., Rips, L.J., Rasinski, K. (2000), *The Psychology of Survey Response*, Cambridge: Cambridge University Press.

Vick, S. and Weidman, L. (1989), 'Reporting on Income Recipiency by Self and Proxy Respondents in SIPP', *Proceedings of the American Statistical Association*, Section of Survey Research Methods, Washington, DC, pp. 315-319.

Watson, I. (2008), 'Low Paid Jobs and Unemployment: Churning in the Australian Labour Market, 2001 to 2006', *Australian Journal of Labour Economics*, Vol. 11, No. 1, pp. 71–96.

Watson, N. and Wooden, M. (2004), 'The HILDA Survey four years on', *Australian Economic Review*, Vol. 37, pp. 343-349.

Weidman, L. (1986), 'Investigation of Gross Changes in Income Recipiency from the Survey of Income and Program Participation', Proceedings of the American Statistical Association, Section of Survey Research Methods, Washington, DC, pp. 231-236.