SEAM EFFECTS CHANGES DUE TO MODIFICATIONS IN QUESTION WORDING
AND DATA COLLECTION STRATEGIES
A COMPARISON OF CONVENTIONAL QUESTIONNAIRE AND EVENT
HISTORY CALENDAR SEAM EFFECTS IN THE PSID

By
Mario Callegaro

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Doctor of Philosophy

Major: Survey Research and Methodology

Under the Supervision of Professor Robert F. Belli

Lincoln, Nebraska
May, 2007
A seam effect occurs in panel studies when within-wave changes are less frequent than between-wave changes (comparing data gathered from two different interviews). This study explores the changes in the magnitude of seam effects among labor force states (employment, unemployment, not in labor force) using the last seven waves of the Panel Study of Income Dynamics collected between 1995 and 2005. The panel underwent several changes: data were collected with conventional questionnaires (CQ) until 2001. The interval between waves was changed from one year to two years in 1997. The data regarding labor force transitions were collected with Event History Calendar (EHC) instruments starting in 2003. The questionnaire was also changed: one modification took place when implementing the two year recall period and the second when starting data collection with EHC.

Data collected with EHCs show a decrease of seam effects in comparison to the previous waves collected with CQ on a two year recall period. A new undocumented phenomenon was also found in the data. When implementing the two year recall period in the CQ waves, a within-wave seam effect appeared, that is a seam effect between the first year and the second year of the two-year reference period reported in the same interview. This effect disappeared in EHC interviews and is most likely due to a change in the questionnaire design. The estimates of labor status changes most affected by seam bias were the transitions from “employed” to “not in the labor force”, and from “not in
the labor force” to “employed”, regardless of the data collection methodology or the length of the recall period. Lastly, the magnitude of one year recall period seam bias was lower than any two year recall period seam points. The results are discussed in light of theories of seam effects, questionnaire design, and with references to the literature on EHC and CQ interviewing methods.
Use of panel design introduces new kinds of measurement error. Response variance, which does not lead to bias in cross-sectional estimates for a single variable, can cause substantial bias in the kinds of longitudinal estimates collected with panel data.

(Slightly adapted from Kalton 1998, p. 63)
Acknowledgements

First of all I would like to thank my supervisor, Robert Belli who supported this study at every step and did more than any student should expect. I wish every PhD candidate an advisor like Bob, in that case we would have less students in need of “dissertation support groups” organized at major universities in the country. Allan McCutcheon recruited me for the master program, supported the decision to undertake the doctoral program, and was a great mentor during these years. I owe him my career change and consequently, this PhD. My dissertation committee (Brian Bornstein, Dan Hoyt, Jon Krosnick, and Allan McCutcheon) stimulated and reinforced the initial dissertation idea opening new research paths that I originally did not envision. I see them basically as coauthors.

This dissertation could not have been written without financial support from the Panel Study of Income Dynamics which sponsored my last year of PhD. The PSID also funded many visits to the PSID data center, University of Michigan, Ann Arbor. They provided me with a quiet work environment in which I was able to concentrate and make fast progress in the data management section of the dissertation. Part of their data not in the public release was made available by the PSID, in particular the pre-release data for the 2003 and 2005 wave. The author wishes to thank Kate McGonagle, Robert Schoeni and Frank Stafford for having believed in my research hypothesis and having facilitated the completion of this dissertation in any way possible.

The Gallup Organization sponsors the Survey Research and Methodology program. Their generosity enabled me and other students to obtain master and PhD degrees in survey methodology at the University of Nebraska, Lincoln. I owe them a lot.
Many colleagues made comments and suggestions, in particular Trent Buskirk, Fred Conrad, Robert Groves, Daniel Hill, John Fitzgerald, Annette Jäckle, Jim Lepkowski, Jeffrey Moore, Peter Lynn, Robert Nielsen, Lance Rips and Norbert Schwarz. I am solely responsible for the analyses and interpretations presented here.

Tecla Loup’s assistance was invaluable in handling the dataset and the pre-release waves. Without her help in navigating the PSID datasets this dissertation would have been taking a much longer time and possibly I would have made some mistakes in the data management.

The main findings were presented at the Methodology of Longitudinal Surveys conference at the University of Essex, UK in July 2006. The conference had a strong impact on this dissertation from an intellectual point of view. I thank the conference organizers for that.

Other people helped in different ways. Alexandra Achen double checked the results of the last two waves of the panel. Femke De Keulenaer taught me most of what I know about panel surveys by reading and discussing her dissertation chapters and gave me advice on how to proceed in the data analysis. Amanda Shields of the U.S. Census Bureau kindly scanned and made available online some seam effects papers of the Survey of Income and Program Participation.

My family always supported and encouraged me in this project, being very understanding when I had to postpone my super optimistic plans to graduate. Lastly, but not of less importance, Ana Villar was able to keep up with me during the final stressful and sometime frustrating moments of the dissertation.

Mario Callegaro April 2007
# Table of contents

**Introduction**............................................................................................................................... 2

**Chapter 1. Seam effects in longitudinal surveys** ................................................................. 6

1.1 Defining seam effects .................................................................................................. 7

1.1.1 *Historical note* .......................................................................................................... 9

1.1.2 *Why seam effects are a problem: magnitude studies* ............................................ 9

1.1.3 *How seam effects are illustrated and the possible combinations of transitions that create seam effects* .................................................................................. 10

1.1.4 *Record check studies* ........................................................................................... 12

1.2 Explanations for seam effects ..................................................................................... 15

1.3 Factors affecting the magnitude of seam effects ....................................................... 18

1.4 Data collection strategies proven effective in reducing seam effects ................... 21

1.4.1 *Dependent interviewing* ....................................................................................... 22

1.4.2 *Calendar aided interviewing* ................................................................................. 25

1.4.3 *Keeping the same interviewer at both waves* ...................................................... 27

1.4.4 *Manipulation of question wording* ....................................................................... 28

1.5 Seam effects in labor force transitions ......................................................................... 29

1.6 Conclusions from the literature ................................................................................... 31

**Chapter 2. The Event History Calendar data collection methodology and its application in the PSID** ................................................................................................................. 33

2.1 A definition of the Event History Calendar ................................................................ 34

2.2 Quality of EHC data .................................................................................................... 39

2.3 EHC and the structure of autobiographical memory .................................................. 42

2.4 The PSID computerized EHC .................................................................................... 44

2.4.1 *EHC techniques* .................................................................................................... 47

**Chapter 3. The Panel Study of Income Dynamics (PSID) data collection procedures** ................................................................................................................................. 49

3.1 The PSID: the first household panel in history ........................................................... 50
3.1.1 Sample design ........................................................................................................ 50
3.1.2 Definitions of eligible sample members ............................................................... 51
3.1.3 Following rules ..................................................................................................... 52

3.2 Data collection ........................................................................................................... 55
3.2.1 Response rates ...................................................................................................... 57
3.2.2 Panel maintenance, tracing and tracking rules ..................................................... 60
3.2.3 Field procedures and interviewing rules ............................................................... 60
3.2.4 Interviewer training and monitoring ................................................................. 61

Chapter 4. Labor force transitions and seam effects in the PSID ......................... 64
4. 1 Study’s hypotheses and data analysis plan .............................................................. 65
4.1.1 Main hypothesis .................................................................................................. 65
4.1.2 Corollary hypotheses ......................................................................................... 67
4.1.3 Expected findings .............................................................................................. 68
4.1.4 Data analysis plan ............................................................................................. 69
4. 2 Data and methods .................................................................................................. 70
4.2.1 Question wording for labor force status ............................................................. 72
4.2.2 Variables ............................................................................................................ 76
4.2.3 Self and proxy answers: a possible confounding factor ...................................... 79
4.2.4 Data treatment, imputation and issues with some variables .............................. 83
4.3 Primary findings ........................................................................................................ 85
4.3.1 Comparison of seam points with previous November-December status changes 88
4.3.2 Comparison of CQ and EHC magnitude of seam points ..................................... 89
4.3.3 Comparison of one year recall period seam magnitude and two year recall period seam magnitude .......................................................... 90
4.3.4 Comparison of two year recall period seam points ........................................... 91
4.3.5 An unexpected phenomena: within-wave seam bias .......................................... 92
4.3.6 Labor force status changes with the highest seam bias ...................................... 94
4.4 Secondary findings .................................................................................................. 95
Index of Tables

Table 1.1 Possible Combinations of Mutually Exclusive States in Labor Force Status... 29
Table 3.1 PSID Following Rules Examples for Special Cases...................................... 54
Table 3.2 Response Rates and Sample Sizes for the 2003-2005 Waves ...................... 57
Table 4.1 Examples of EHC Conversions from Thirds of a Month to Month String for
Labor Force Status ........................................................................................................ 77
Table 4.2 Examples of Possible CQ Status Changes.................................................. 78
Table 4.3 Test of Marginal Homogeneity to Assess Differences in the Mixture of Self or
Proxy Answers Among Different Seam Points ............................................................ 81
Table 4.4 Pearson $r$ Correlations Between Self and Proxy Answers from the Same
Household ......................................................................................................................... 82
Table 4.5 Number of the Nine Classifiable Status Changes and Percentage of
Unclassifiable Status Changes for Each Reference Period and at each Seam Point ...... 85
Table 4.6 Test of Marginal Homogeneity Between Seam Points and Immediate Previous
Status Changes November-December ........................................................................... 89
Table 4.7 Test of Marginal Homogeneity Between CQ Seam and EHC Seam.......... 89
Table 4.8 Chi Square Test Between One Year Seam Point and Two Year Seam Points. 91
Table 4.9 Test of Marginal Homogeneity Between Two Year Recall Period Seam Points
........................................................................................................................................ 92
Table 4.10 Test of Marginal Homogeneity Between Within Wave Seam Points and
Immediate Previous Transitions November-December................................................ 93
Table 4.11 Percentage of Movers Spurious Transitions at the Seam When Compared to
the November to December Transitions ....................................................................... 96
Table 4.12 Predictor Variables Used in the Logistic Regression, their Categories and
Measurement Level....................................................................................................... 98
Table 4.13 Dummy Coding of the Predictor Variable Self-Proxy Status..................... 99
Table 4.14a CQ 1998_99 Logistic Regression Output ............................................... 102
Table 4.14b EHC 2002_03 Logistic Regression Output .............................................. 102
Index of Figures

Figure 1.1 1983-84 Survey of Income and Program Participation Month to Month Transition Rates ................................................................................................................................. 8
Figure 1.2 Possible Cases Resulting in Artificial Transitions that Increase the Between Wave Transition Rate ................................................................................................................. 11
Figure 1.3 Possible Case Where a True Change Does Not Create a Seam Effect........... 12
Figure 1.4 Simplified Interview Schedule for a Hypothetical Panel with Reference Period of 4 Months ................................................................................................................................. 17
Figure 1.5 Unemployment Compensation Transitions by Month. PSID and SIPP Data 1983-84 ........................................................................................................................................... 21
Figure 1.6 Calendar Shown to the Respondents During the Second Wave of the SIPP Calendar Experiment ........................................................................................................................................... 26
Figure 1.7 Ratio of Average Monthly Seam to non Seam Transition for two Conditions: Calendar Experiment and Regular SIPP data ................................................................................................................................. 27
Figure 1.8 Labor Force Transition Rates for the SIPP ........................................................................................................................................................................................................... 30
Figure 2.1 Reproduction of the Instrument used in the 1997 PSID Paper and Pencil Calendar Method Study ........................................................................................................................................................................................................... 37
Figure 2.2 Calendar used in the Engel et al. (2001) study of Migrant Farm Workers ..... 40
Figure 2.3 Schematic PSID Questionnaire ....................................................................... 44
Figure 2.4 Landmark Events Screenshot .......................................................................... 45
Figure 2.5 Residential Domain Screenshot ....................................................................... 46
Figure 3.1 PSID Major Design Changes Timeline ........................................................... 56
Figure 3.2 PSID Schematic from McGonagle & Schoeni .............................................. 59
Figure 4.1 Combinations of Head and Wife Answers at the Seam Creating Four Self/ Proxy Reports ........................................................................................................................................................................................................... 67
Figure 4.2 PSID Waves Used in the Analysis and Seam Points ........................................ 71
Figure 4.3 CQ Questionnaire Flowchart for Labor Force Variables 1995-97 and 1999, 2001 T-1 ........................................................................................................................................................................................................... 73
Figure 4.4 CQ Questionnaire Flowchart for Labor Force Variables 1999 and 2001 $T - 2$ ................................................................. 74

Figure 4.5 Questionnaire Flowchart for EHC Labor Force Variables 2003 and 2005 $T - I$, $T - 2$ .................................................................................................................. 75

Figure 4.6 Mixture of the Four Combinations of Self or Proxy at each Seam, Percentage of Answers .......................................................................................................................... 80

Figure 4.7 Month to Month Status Changes of Movers in Waves 1995 – 2005, All Answers, no immigrant Samples. .................................................. 87

Figure 4.8. Month to Month Transitions of EN and NE Movers in Waves 1995 – 2005, All Answers, no Immigrant Samples .................................................. 94
Citation note

Please send me an email if you want to cite this dissertation. In this way I can provide you with the most recent citation and current updates on the status of this work.

I can be reached at: mario.callegaro@gmail.com

Thanks
Introduction
Seam effects are typical phenomena observed in longitudinal data collection. A seam effect occurs in longitudinal studies when within wave changes are less frequent than between wave changes (with data coming from two different interviews/waves). In other words, a seam effect is present when month-to-month changes in responses are much larger for the seam months than for adjacent months away from the seam.

Labor force transition rates are not immune to seam effect bias when analyzing labor market dynamics in panel studies (Paull, 2002). Econometricians have found that when Labor Force Surveys (LFS) collected with a panel design are used to study labor market dynamics, often times reported changes in status tend to cluster at the seam at a significantly higher rate than within the wave\(^1\). The importance of this finding stems from the practical consequences that the bias will have in the way data are used; the information of transitions at the seam cannot be interpreted as it is.

Up to now seam effects have been observed in panel data collected using a standardized conventional questionnaire (CQ). Although some interviewing strategies such as dependent interviewing, for example, have been shown effective in reducing seam effects, Event History Calendar (EHC) data collection methodology has not been tested yet on its performance for the seam effect problem.

This study takes advantage of a change in data collection strategies of the Panel Study of Income Dynamics (PSID). The PSID collected data using conventional questionnaires until the 2001 wave. Starting in 2003, sections of the questionnaire, including labor force questions, were collected with the EHC. Due to this change, now two waves of EHC (2003 and 2005) data about labor force history are available for analysis enabling to study the magnitude of the transition rates at the seam for the 2003 – 2005 waves. The magnitude of the EHC seam can now be compared with the magnitude of the seam for the previous waves of the PSID collected with CQ interviews. From an experimental de-

---

\(^1\) Not all LFS show seam effects. For example, the Current Population Survey (a rotating panel design) conducted in U.S. asks respondents to report their job status only referring to the calendar week before the interview and not for the entire reference period, that is of one month (Current Population Survey, 2002). Because of this design, there are no within-wave data thus precluding the computation of possible seam effects.
sign point of view, this is a before-after within subject experiment, where most of the same panel participants provided answers replying to a CATI Conventional Questionnaire, and at later waves, to a computerized telephone EHC interview.

The starting point of the dissertation is the theoretical framework of the EHC data collection methodology that has been shown to be superior to CQs in collecting retrospective data in terms of the reduction of underreporting, and in terms of precision of the location of the events in time (Belli, Shay, & Stafford, 2001). The general hypothesis is that panel data collected with EHCs should show a decreased seam effect in comparison to the same data collected with CQs.

The dissertation is organized in five chapters. The first chapter presents an overview of the nascent literature about seam effects. In fact the first paper presenting findings about seam effect dates back to 1983 (Czajka, 1983). After defining what is a seam effect and reporting validation studies measuring its impact on data quality, current explanation for seam effects are discussed. Respondents’ characteristics and data collection features have shown to have an impact on the magnitude of seam effects. Few studies with validation data have been able to measure those characteristics. Some data collection strategies are shown to be effective in reducing seam effects, the most successful being dependent interviewing. The chapter ends with a review of the studies about seam effects in labor force transitions.

In the second chapter the Event History Calendar methodology is defined. After presenting an overview of the technique, studies measuring the quality of the data collected with EHCs in comparison to CQs are discussed. Parallels between the EHC instrument and the structure of the autobiographical memory explain why the EHC elicits better recall and provides more precision in the location of events in time. Lastly, the PSID EHC computerized instrument is explained, together with information about interviewer training and how the interview is actually conducted.

In the third chapter, the Panel Study of Income Dynamics design and data collection strategies are laid out to enable the reader to better understand the data analysis and possible confounding factors presented later on. Particular emphasis is placed on the last
seven waves of the panel that will be used in the analysis because the PSID has gone through major design changes that are hypothesized to have an impact on the seam bias.

The fourth chapter starts with the general and the specific research questions of this dissertation. Based on the results from the literature, I expect to find a reduction of a seam effect in the EHC. The data analysis plan is then explained together with caveats and possible confounding factors that will be taken into account. Main findings and secondary findings close the chapter.

The fifth chapter concludes the dissertation. A summary of the findings and a possible strategy to further tackle seam effects in the PSID is proposed. Lastly, a future research agenda on seam effects is delineated in order to better understand a phenomenon that is not fully explained yet.
Chapter 1.

Seam effects in longitudinal surveys
1.1 Defining seam effects

A seam effect occurs in longitudinal studies when within wave changes are less frequent than between wave changes (with data coming from two different interviews/waves). For example, a seam effect is present when month-to-month changes in responses are much larger for the seam months than for adjacent months away from the seam (Rips, Conrad, & Fricker, 2003). In the literature, a seam effect is sometimes referred to as a “heaping effect” (Kraus & Steiner, 1998; Torelli & Trivellato, 1993b) and some authors classify it in the category of panel conditioning (Burkhead & Coder, 1985; Ruspini, 2002). Panel conditioning refers to the effect that panel respondents’ answers can be affected by their participation in previous waves (Kalton & Citro, 1993).

Almost all seam effect charts follow a typical pattern, as shown in Figure 1.1. In the figure, month to month transition rates were computed for each pair. The plot shows the change in status from not receiving Social Security benefits to receiving them, and from not receiving Food Stamps to receiving them. The first transition (1st to 2nd month) is used as a baseline with an index of 100.0. Given that the Survey of Income and Program Participation (SIPP) wave length is of 4 months, the seam between waves occurs from the 4th to the 5th month and from the 8th to the 9th month. The reader should note that at the seam junctions the transition rates are almost three times higher than the average within wave transitions.
Seam effects have been observed in different panels, with different reference periods, and in many countries. A few examples include the SIPP (Bassi, 1998; Kalton, 1998; Kalton & Miller, 1991; Marquis, Moore, & Huggins, 1990), the Income Survey Development Program (ISDP) – precursor of the SIPP (Moore & Kasprzyk, 1984), and in the Panel Study of Income Dynamics (PSID) (D. H. Hill, 1987). The Canadian Labour Market Activity Survey (LMAS) (Murray, Michaud, Egan, & Lemaitre, 1991), the Canadian Survey of Labour and Income Dynamics (SLID) (Cotton & Giles, 1998), the British Household Panel Survey (BHPS) (Jäckle & Lynn, 2004) and the German Socio-Economic panel (GSOEP) (Kraus & Steiner, 1998) have shown seam effects as well.
1.1.1 Historical note

The seam effect is a quite recent issue in survey research. The problem was noticed for the first time by Czajka (1983) in a reanalysis of the linked first and second wave data of the ISDP used in a previous study by Lepkowski and Kalton (1981). Czajka noticed that the turnover of receipts from various income types was more pronounced between waves than within waves. The next year Moore and Kaspryk (1984) did the first seam effect analysis of the ISDP finding seam effects for a number of variables. At the same time Bowers and Horvath (1984) completed an analysis of two linked Current Population Survey waves and found that: “on average, the between-monthly surveys’ change in the duration of unemployment among persons in a continuous spell exceeded the actual elapsed time” (p. 148). The term “seam” was introduced by Burkhead and Coder (1985, p. 353) in their analysis of the 1983-84 SIPP data. Until 1991, seam effect papers were found in conference proceedings only, specifically in the American Statistical Association’s conference proceedings. It was in that year that the first peer reviewed paper using the term “seam effect” was published in the Journal of Official Statistics (Kalton & Miller, 1991).

1.1.2 Why seam effects are a problem: magnitude studies

Seam effects are a problem in the attempt to collect accurate survey estimates. Magnitude studies show how the scale of seam effects is large enough to be considered a problem and not a minor source of noise in the data distribution (Willis, 2001).

For example, data from the first and second waves of SIPP showed that the between wave recipient turnover rate for income coming from wage and salary was 9% compared to a within-wave average of 3%. The on-seam turnover rate for income coming from food stamps was 1.47% compared to the average off-seam rate of 0.48% (Moore & Kasprzyk, 1984). The on-seam unemployment compensation transition rate was 48% compared to an average off-seam rate of 16% in the PSID (calculated from D. H. Hill, 1987). In the Canadian LMAS panel, 180,000 people shifted their job search approach
from “Did not want a job” to “Wanted a job” at the seam, compared to an average month
to month number of 21,450 from the previous wave (calculated from Lemaître, 1992).

Interestingly enough, beside few exceptions (Kalton & Miller, 1991; Moore &
Kasprzyk, 1984) the studies mentioned above rarely performed any statistical test to
compare the on-seam data with the off-seam data. One reason might be of the “non ran-
dom” pattern of the data distribution (a peak at every seam). It is however true that for
some variables the difference between on-seam and off-seam data does not reach statisti-
cal significance, as demonstrated by Moore and Kasprzyk (1984).

1.1.3 How seam effects are illustrated and the possible combinations of transitions that create seam effects

There are mainly two ways to illustrate seam effects. One way is to compare the transi-
tion between two waves (on-seam transition) with the month to month transitions within
each wave (off-seam transitions). This will result in a distribution such as the one in Fig-
ure 1.1. Other authors compare on-seam transitions with the average of off-seam transi-
tions.

In the following figures, possible combinations of transitions that create real or
spurious transitions are reviewed. Figure 1.2 shows a scenario where the true status of
transitions is compared with some possible answers that create seam effects.
Figure 1.2 Possible Cases Resulting in Artificial Transitions that Increase the Between Wave Transition Rate. The “1” and the “2” Indicates Two Different Statuses

In figure 1.2a there is no change at the seam, but there is one change within the first wave, precisely between the months of November and December. The forgotten change in December creates a false transition when linking the two waves. Other possible causes for the spurious transition shown in figure 1.2a are interviewer or coder error. In figure 1.2b although there is no change at the seam, a change is reported in January. Possible causes are reporting erroneous information or interpreting the question differently at wave T + 1. In figure 1.2c and 1.2d effects of misplacing a change in time create seam effects. In the first case (1.2c), the transition from October to November is reported from December to January due to forward telescoping (reporting events as happening more recently than they actually happened). In the second case (1.2d), a transition that happened in wave T + 1 from January to February is pushed back in time of one month thus creating a false transition between December and January. This is a case of backward telescoping (reporting events as happening more at a more distant time than they actually occurred).

Figure 1.3 portrays a case where the true status is not correctly reported but does not create an increase of spurious transitions at the seam because the forgotten change happens within the wave. This possibility was contemplated by Martini & Ryscavage.
(1991). The point I want to make in showing Figure 1.3 is that seam effect does not reflect all the reporting errors that exist.

<table>
<thead>
<tr>
<th>Month</th>
<th>S</th>
<th>O</th>
<th>N</th>
<th>D</th>
<th>J</th>
<th>F</th>
<th>M</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>True status</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reported status</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1.3 Possible Case Where a True Change Does Not Create a Seam Effect

Before discussing the different explanations for seam effects that have been hypothesized and tested in the literature, it is useful to review the few record check studies that assessed the magnitude and the direction of the errors. Doing this review will facilitate the evaluation of the seam effect explanations.

1.1.4 Record check studies

Goudreau, Oberheu and Vaughan (1984) conducted the first record check study of the ISDP, the precursor of the SIPP panel. The variable of interest was “Aid to Families with Dependent Children” (AFDC). 260 families were sampled in five different states from the payment files maintained by the welfare agency in each state. The interviews were conducted in July and asked about AFDC benefits received during April, May, and June. Data collected from these interviews were then compared to the administrative records provided by the state welfare agencies. One finding was that 9% of persons underreported having received AFDC and 13% misclassified AFDC as another type of cash assistance. Those who failed to report having received AFDC assistance were likely to have received it for a part of the reference period of the survey. Even more interesting for our purposes, the most common error in reporting the amount of income received was the tendency to report the amount of most recent payment as the same for every previous month of the reference period, when in reality that amount varied. This finding gives support to the constant wave response theory (Young, 1989) discussed later.
Marquis and Moore (1989) conducted a record check study for the first two SIPP interviews of 1983 and 1984. The SIPP panel collected data every four months. The variables selected were participation in the following programs: Aid to Families with Dependent Children (AFDC), Food Stamps (FOOD), Supplemental Social Security (SSI), Unemployment Insurance (UNEMP), Workers Compensation (WORK), Civil Service Retirement (CSRET), Social Security (OASDI), and Veteran Benefits (VETS). Administrative records from 4 different states were matched with the reports from the SIPP interviews conducted in 1983 and 1984. The authors computed a transition bias that is the sum of the differences between panel data and administrative records divided by the true number of transitions and multiplied by 100. The transition bias was negative for the off-seam months with an average of -38% across the programs and positive for the on-seam months with an average of +802% across the programs. In other words, the seam effect is the result of too many transitions measured at the seam (over-reporting) and too few measured elsewhere (underreporting). In a subsequent reanalysis of the same data, some causes of response error were tested: forgetting (underreporting), memory decay, and external telescoping (Marquis, Moore, & Huggins, 1990). The forgetting hypothesis was not supported by the data. In fact, although the record check study showed a predominance of underreporting, some variables showed overreporting as well. The memory decay hypothesis (more underreporting for recall of “4 months ago” compared to the immediately previous month) was falsified by the data. For most programs underreporting rates were the same for recent and past events. Finally, the level of overreporting in wave 1 (unbounded interview) was the same as in wave 2 (bounded interview) making the authors conclude that external telescoping was not responsible for response error.

Similar results were obtained some years later in the SIPP Cognitive Research Evaluation experiment (Moore, Marquis, & Bogen, 1996). A small scale two wave panel was created with the same design of SIPP and interviews were conducted in 1992 and 1993. In order to conduct the experiment, part of the sample (600 cases for each program) was drawn directly from AFDC, Food stamps, SSI and UNEMP participants. One reason for this selection strategy was to have true program participants. The second reason was
of a more practical nature. Since program participants are rare in the general population, it would be more efficient to draw a sample from the general population and then find the program participants, that to do the opposite. The record check study was combined with an experiment. In the treatment group the respondents were encouraged to use personal income records. Two interviewing team were used, one for each condition. In the treatment group more emphasis was given into training interviewers to obtain accurate answers and trying to do the interview in a distraction-free environment. In the control group interviews were done using the standard SIPP protocol. The results of the experiment showed mixed outcomes. Although the experimental condition canceled the off-seam bias (underreporting), it exacerbated the on-seam bias, which is more overreporting happened at the seam (+106% vs. +43% in the control group, p. 40). The authors did not offer an explanation for this effect.

The PSID Validation Study (PSIDVS) was conducted in a manner similar to the SIPP Cognitive Research Evaluation experiment in the sense that a PSID-like two wave panel was created ad hoc for the experiment. A two wave panel was purposely created using the PSID employment questionnaire (Bound, Brown, Duncan, & Rodgers, 1994). The goal was to evaluate the quality of employment questions from the PSID standard core questionnaire. One large manufacturing firm with several thousand employees was selected to conduct the validation study. The first wave of data was collected in 1983, while the second took place in 1987. The company record data were used to validate the two waves’ answers. The papers written about the PSIDVS do not focus directly on seam effects and do not attempt to measure them. Nevertheless, two important findings in terms of measurement error should be mentioned. Workers with lower-than-average earnings tended to over report their usual pay period earnings while workers with higher-than-average earnings tended to underreport it (Rodgers, Brown, & Duncan, 1993). The second finding is that the percentage of the actual unemployment spells declined with the lengthening of the recall period, going from 49% for eight or less months to 25% for 19 or more months (Bound, Brown, Duncan, & Rodgers, 1990). This latter finding gives support for the memory decay explanation discussed in the following subsection.
1.2 Explanations for seam effects

There are many possible causes that are believed to be responsible for the seam effect phenomena. They can be classified into four categories: data management issues, forms of satisficing, memory issues, and respondent errors. These four classes of causes frequently occur together and the contribution of each depends on the kind of variable studied.

Burkhead and Coder (1985) identify data keying error as a possible source for seam effects in the SIPP panels. Unpublished work done on wave 1 and 2 of the BHPS reported by Halpin (1998) showed that for respondents who were employees in both waves, and had not changed their jobs, 32% changed their 3 digit Standard Occupation Classification (SOC) code and almost 18% changed their SOC major group. An analysis of a sample of the original paper questionnaire revealed that the bulk of cases were caused by either different coding of what essentially was the same job description, or different descriptions of the same job. Similar results were found in the SIPP by Kalton, McMillen and Kasprzyk (1986). Halpin (1998) shows how inconsistencies can be created by merging the first five waves of the BHPS into a unique dataset. The reconciliation of the waves creates different percentages of matching among the variables of interest, in this case employment status, depending on the variables used as keys to merge the files. Mismatches and edit issues were hypothesized previously by Moore and Kasprzyk (1984) for the ISDP panel, and by Cotton and Giles (1998) for the SLID. Lastly, it is possible that procedures for assigning values for missing items (Lynn, Buck, Burton, Jäckle, & Laurie, 2005) and interviewer errors are contributing to seam effects patterns (Burkhead & Coder, 1985).

Burkhead and Coder (1985) talk about respondents purposively reporting no change for some statuses in order to shorten the length of the interview. The same idea is discussed by Martini (1988), who debates the possibility for respondents to omit details that could lead to further questioning. Young (1989) and Martini (1989) introduce the term “constant wave response” to define respondents who give an answer for earlier
months in an interview period identical to the other months that are asked about within the same interview. Young (1989) found constant wave response effects for both continuous and categorical variables (e.g. yes no). Czajka (1999) defines the same phenomena in different terms and talks about respondents treating the entire reference period as if it were a single observation period. Kalton and Miller (1991) discuss the possibility that respondents repeat the same answer because it simply requires less effort. All these findings are now called within-wave underreporting, (Biemer & Lyberg, 2003, p. 139) a phenomena that falls under the concept of satisficing (Krosnick, Narayan, & Smith, 1996).

Forms of satisficing can be exacerbated by the design of the questionnaire. Because panel respondents answer the same questions each wave, they can learn that a ‘Yes” response often leads to more detailed questions. In this case a design with filter items where the answer “yes” opens up to further questions, can “invite” the respondents to satisfice and just say “no” (Hoogendoorn, 2004).

Another way of thinking about seam effects is to view them as the contrast between memory for the most recent portion of the response period of the earlier wave, and memories for the most remote portion of the response period of the later wave. The latter are likely to be an estimate more than a direct recall (Rips, Conrad, & Fricker, 2003). Figure 1. 4 exemplifies this concept.
Figure 1.4 Simplified Interview Schedule for a Hypothetical Panel with Reference Period of 4 Months

Note: The dotted curves highlight the fact that recalls of events that happened 4 months before the second interview are compared with reports of events that happened 1 month before the first interview.

Rips, Conrad and Fricker (2003) observe simply that recall accuracy decreases over time. This loss of recall accuracy leads to underreporting of remote changes in the reference period, thus creating spurious changes across the seam. The best evidence of this concept comes from Kalton and Miller (1991) who analyze the Social Security benefits in the SIPP. Taking advantage of the SIPP design that rotates the groups of respondents being interviewed at different points in time, the authors showed evidence for the memory decay hypothesis.

Seam effects can be the result of respondents reporting erroneous information that leads to the misclassification of a spell (Martini, 1989; Martini & Ryscavage, 1991). For example, a period of layoff can be reported as a period of unpaid vacation. For each misclassified spell, a pair of transitions is recorded incorrectly. Martini and Ryscavage (1991) also make the distinction among two different situations: the “true” spell is com-
pletely contained in the reference period, and the “true” spell spans across two or more reference periods and only the portion in one reference period is misclassified. In the first case no seam effect occurs because of the misclassification, but in the second case the misclassification contributes to the seam effect.

Multiple causes produce seam effects. There is no single study that has been able to measure the specific contribution of each cause. A further complication comes from the fact that the magnitude of seam effects changes for different variables (Young, 1989). For example classification errors are more likely to happen for variables such as job description (coder inconsistencies) or job status (respondent error). Other variables such as receiving Social Security Income (“yes”/”no” answer) have no possibility of coder error since there is no coder involved.

Constant wave response has its counterpart in the autobiographical memory literature. A similar phenomena is called retrospective bias (Ross, 1989). Our memory for earlier periods of our lives lacks details. In order to reconstruct them, we may extrapolate current characteristics and apply them to the past. On one hand, if we expect that a characteristic did not change much, we may infer that its value is identical to the present one. On the other hand, if we expect that a characteristic changed quite a lot, we may exaggerate the amount of difference between the past and the present. In the panel case, respondents may use their current status (easier to remember) to estimate earlier values, but fail to adjust for changes in the more remote portions of the reference period.

1.3 Factors affecting the magnitude of seam effects

Different factors affect the magnitude of seam effects. They can be classified into two categories: respondent characteristics and design level characteristics.

There are mixed findings when analyzing self-reported answers versus proxy answers for seam effects. The key to coming to a conclusion lies in the kind of variables analyzed and, most importantly, the relation between the variables of interest and the respondent selection (Vick & Weidman, 1989). For example, when studying employment
status, a proxy can be used because the reference person is working and thus less available to the interview, creating a self-selection problem. Self-selection is one of the methodological shortcomings of much of the research which aims to address the self/proxy issue (Moore, 1988). Moore and Kasprzyk (1984) found that a proxy-proxy report (first wave-second wave) and a proxy-self and self-proxy report was four times higher at the seam in comparison to the off-seam than a self-self report that was 2.5 times higher at the seam for variables such as income wage and salary income. For income coming from self employment the proxy-proxy report was 12 times higher at the seam and the self-proxy/proxy-self was 16 times higher at the seam in comparison to the self-self report that was only 8 times higher at the seam. They did not find effects of such magnitude for other variables such as Social Security benefits or AFDC. A possible explanation is that the recipiency of Social Security benefits and food stamps is common knowledge in the household, much more than the precise transitions of the respondent in his/her job status. Vick and Weidman (1989) do not find support that the use of proxy respondents increases the number of transitions at the seam. Similar results come from the Canadian LMAS panel (Murray, Michaud, Egan, & Lemaître, 1991). In summary, a proxy respondent is not necessarily less knowledgeable than the reference person, depending on the topic and the amount of discussion on it by the proxy and the other members of the households (Sudman, Bradburn, & Schwarz, 1996).

Some respondents’ characteristics were found to amplify inconsistencies for between-wave reports even if few studies are available. Taking advantage of reinterview data, Hill (1987) was able to create a variable at the individual level where consistent reports were treated as true transitions and inconsistent reports were treated as spurious transitions. Consistent reports were defined when the answer about employment, unemployment, and not in the labor force for the months of December 1983 and January 1984 were the same of the first interview done in 1984 and of the second interview done in 1985. He was then able to create a multinomial logistic model with race, age, sex, education, income, and months with the current employer as independent variables. For the dependent variable employment status, African Americans), older respondents, and having
worked few month with the current employer were found to amplify inconsistencies thus creating spurious transitions at the seam. Using validation data Jäckle and Lynn (2004) found that older people were more likely than others to provide erroneous reports or missing data, thus creating spurious transitions between waves.

Since the first studies of seam effects it has been noted that the magnitude depends on the types of variables analyzed (Burkhead & Coder, 1985; Kominski, 1990). Not surprisingly, less memorable events tend to be forgotten more, as is the case of less memorable jobs, e.g. short duration and without a pension plan (Murray, Michaud, Egan, & Lemaître, 1991).

Lastly, the length of the reference period increases the magnitude of seam effects. Memory decay and forgetting play a role especially if the reference period increases. For example Hill (1987) studied the magnitude of seam effects for variables such as unemployment compensation and food stamps recipiency in the SIPP (quarterly data collection) and in the PSID (yearly data collection) panel. Taking advantage of the fact that the two studies overlapped for some months and of a very similar question wording, Hill showed how the magnitude of seam effect was higher for the PSID in comparison to the SIPP. Figure 1.5 visualizes Hill’s findings.
There are two main ways to reduce seam effects: one is with a panel design that includes an overlap of recall periods between waves, the second is using special data collection strategies.

The Australian HILDA panel has an overlap of six months between each wave of the panel (Watson & Wooden, 2004) for events regard education and job history. For this reason it is possible to obtain two measurements at the seam of the two waves (test-retest). If data are found to be inconsistent between the two data collection, the reported closest to the interview were used at the seam. For example 27% of people gave inconsistent answers between the two waves regarding their unemployment status (Carroll, 2006).
Although this method does not completely solve the seam bias issue, it provides a measure of its magnitude.

In the second stream of research, involving special data collection strategies, solutions proposed include the use of dependent interviewing, the use of a calendar, having the same interviewer, and the manipulation of question wording.

1.4.1 Dependent interviewing

With the advent of Computer Assisted Interviewing (CAI) many panels have introduced dependent interviewing as a way to reduce measurement error in general (A. Brown, Hale, & Michaud, 1998). Dependent interviewing was used before computerization as well. Its precursor can be traced to the bounded interviews by Neter and Waksberg (1964).

Dependent interviewing questions are classified as proactive and reactive (A. Brown, Hale, & Michaud, 1998; Mathiowetz & McGonagle, 2000). In proactive dependent interviewing the respondents are reminded of responses given in the previous wave. The previous information is used to aid the respondent’s memory and provide a bounded recall before providing a standard independent question. This strategy is called remind-continue. Another strategy, called remind-confirm, consists in asking the respondent to check and confirm previously recorded data. One more option is to ask if there are any changes since the last wave (remind-still) (Jäckle, in press).

With reactive dependent interviewing the information fed forward is used to carry out edit checks during the interview. Reactive dependent interviewing is used as a form of follow-up for an item nonresponse (e.g. don’t know, don’t remember) or to check the consistency with a previous report. In the first case, item nonresponse follow-up, if for example the respondent does not provide an answer to the income question, he/she is provided with the previous report, asking if that value “sounds about right” (Moore, Pascale, Doyle, Chan, & Griffiths, 2004). In the second case, corrective follow-up, consistency checks can be prompted for everybody on a specific question (e.g. to check
verbatim answer to previous report), to clarify reports that are inconsistent with previous reports, or selectively to be prompted only if the current report differ of more of a certain threshold from the previous report (Jäckle, in press).

Proactive dependent interviewing has been studied by Hill (1994) in an experiment of the SIPP for the collection of industry and occupational information with paper and pencil questionnaires. In one condition the entire set of occupation questions was asked and coded for each wave. In the experimental condition these questions were only asked if the respondent reported a new employer or answered the following screening question in the affirmative: “Have your main activities or duties changed with this employer during the past 8 months?” If no changes were reported, the prior data were brought forward. Hill found a reduction of the rate of change for industry and occupational information with proactive dependent interviewing, and although the technique may have been resulted in an underreporting of a change, “most of the changes missed by the dependent method of collection methodology are noise” (p. 379).

The Canadian Survey of Labour and Income Dynamics (SLID) incorporates both proactive and reactive dependent interviewing (Dibbs, Hale, Loverock, & Michaud, 1995). Reactive dependent interviewing is triggered based on the previous wave data. For example, in January, at the end of the demographic and labor questions, respondents were asked whether they had received unemployment insurance, social assistance (welfare), or workers’ compensation during the previous year. In May (next wave), if an expected income source was not reported, the computer displayed a probe for the interviewer to ask the respondent about the omission. The probe was: “Based on our January interview, we thought we would get an amount for [type of income]. Did we miss it?” The wording of this probe was studied to avoid giving negative feedbacks to the respondent (Webber, 1994). The dependent interviewing strategy used in the SLID improved the underreporting of many variables and improved the data quality as well (Hale & Michaud, 1995).

---

2 This question wording would be now classified as proactive dependent interviewing, remind-still.

3 This question wording would be now classified as reactive dependent interviewing, corrective follow-up
The Survey of Income and Program Participation carried out a program of study about dependent interviewing, the SIPP methods panel project in the late 1990s (Moore, Pascale, Doyle, Chan, & Griffiths, 2004). After many pretests, dependent interviewing was implemented in 2004. When comparing the seam bias of the 2004 waves with 2001, 45 of 42 comparisons showed a statistically significant difference. The percentage of change at the seam was reduced in 2004 for variables of “need-based” public assistance type (e.g. receipt of federal Supplemental Survey Income, AFDC, food stamps) and for non need-based income sources and characteristics variables (e.g. private heath insurance coverage, private pensions, Medicare). Dependent interviewing not only reduced the change at the seam but also increased the change off-seam (Moore, Bates, Pascale, & Okon, in press). This is a very important result confirming the previous SIPP findings from validation data that off seam estimates are generally an underestimate of the phenomena (Marquis & Moore, 1989; Marquis, Moore, & Bogen, 1991).

More recently, an experiment collecting work history data involving independent interviewing, and proactive and reactive dependent interviewing was carried out on a subsample of the UK part of the European Community Household Panel (EHCP) (Jäckle & Lynn, 2004). Due to the small number of cases, independent and reactive dependent interview data were combined. The proactive dependent interviewing clearly reduced seam effects. For example, transition rates among the occupational status were reduced from 32% to 9% in the proactive dependent interviewing group. The authors also sustain that proactive dependent interviewing did not lead to underreporting of change since in the off-seam months the average monthly transition rate was the same for all treatment groups. This finding led the authors to conclude that the interviewing method did not make a difference in the number of transitions within a wave. The main difference between methods lies in what happens at the seam. In further research the above dependent and independent interviewing methods were compared in a validation study of benefit receipt (Lynn, Jäckle, Jenkins, & Sala, 2006). The authors found that dependent interviewing reduces the extent of underreporting of benefit receipt without increasing over reporting. However, some net underreporting still remained with dependent interviewing.
Dependent interviewing have been also used with success with self administered panel surveys (Holmberg, 2004) and with self administered establishment surveys (Hoogendoorn, 2004).

Implementing dependent interviewing poses however some challenges. One is that it adds extra programming time, testing time and overall cost (Hoogendoorn, 2004). Second, if there is a switch of respondent between waves, the information collected from one respondent is going to be fed to another respondent creating a confidentiality issue. One solution adopted by the U.S. Census, for example, is to ask permission to disclose the respondent’s answers to other household members (Chan & Moore, 2006). Dependent interviewing is to date the method that have been proven effective in reducing seam bias in many different panels collected with different strategies.

1.4.2 Calendar aided interviewing

In 1989, a 32-month calendar was used as an aid for SIPP respondents in the Chicago region (Kominski, 1990). After completion of the first face-to-face interview (wave 1), the interviewer filled out (in their own homes) the calendar with the information obtained from the standardized questionnaire. In the second interview (wave 2) the interviewer handed the appropriate calendar to the respondents prior to the start of the interview. During the interview the respondent was able to look at the calendar and the events recorded on it. At the conclusion of wave 2, the interviewers updated the calendar that would be used again in wave 3. Figure 1.6 presents the calendar shown to the respondents in the second interview.
Results from the analysis comparing the ratio of average monthly seam to off-seam transitions with previous SIPP data showed a reduction of the seam transitions for almost all variables analyzed as shown in figure 1.7. Furthermore, examination of the calendar showed that there were numerous instances where the calendar facilitated longitudinal editing and correction of data (Kominski, 1990, p. 556).
Figure 1. Ratio of Average Monthly Seam to non Seam Transition for two Conditions: Calendar Experiment and Regular SIPP data. Data from Kominski (1990)

1.4.3 Keeping the same interviewer at both waves

Vick and Weidman (1989) studied seam effects in the SIPP for income recipiency. They found that keeping the same interviewer for consecutive waves reduced the number of reported transitions at the seam, when compared to having a different interviewer. The interviews were conducted with a face-to-face paper and pencil questionnaire. A possible explanation for the above findings is that the interviewer could have acted as a sort of “dependent interviewing mechanism”, possibly reminding the respondent of previous answers (4 months before) and also facilitating the recall just with his/her presence in the household. Unfortunately this is the only study of this nature thus more evidence is necessary to corroborate the findings.
1.4.4 Manipulation of question wording

Rips, Conrad and Fricker (2003) reproduced a two-wave panel in laboratory settings using a one recall period. During three different experiments they were able to reproduce the higher transitions at the seam compared to the off-seam weeks. The seam effect was a function of the difficulty of the recall task, the length of the recall period, the grouping of questions by item or by reference period (in this case, one week), and the order of retrieval. Seam effects enlarged with an increase of the recall difficulty and length of recall. More importantly, and a new finding from the previous studies, grouping the questions by item increased the constant wave response. In other words if, as it happens for example in the SIPP, a respondent first answers a question about food stamps regarding the last month, and then immediately answers the same question regarding 2, 3, and 4 months before, he/she is more likely to give the same answer for each month. On the other hand, if he/she is asked about food stamps and all the other questions for the last months, and then again the same battery of questions for 2, 3, and 4 months before, it is less likely that the same answers (constant wave response) will be obtained for each month. Asking different questions for a month of the reference period, and then the same group of questions for the month before discourages satisficing. Furthermore, when recall by reference period is coupled with backward order of retrieval (i.e. reverse chronological) it produces less constant wave response, lower seam effects and higher accuracy levels. Even if some limitations have been proposed for this study (Martin, 2001; Willis, 2001) it is indeed the first successful attempt to reproduce seam effects in controlled laboratory settings.

Before moving to the next section is worth mentioning another stream of research that is found in the literature of seam effect. These papers deal with seam effect after the data are collected. Written primarily by statisticians and economists, many strategies are proposed to adjust the data for seam effect. We find, for example, Kalton and colleagues (1992), who propose to allocate a proportion of the spell starts reported in the first reference month to later months in the same interview wave. Another approach is proposed by Bassi (1998) who uses latent class analysis to compensate for the seam effect. Variations
of Cox’ proportional hazard model have been proposed (Petoussis, Gill, & Zeelenberg, 2004) as another strategy to statistically adjust for the heaping in the data. There are many other statistical adjustments that are currently used to deal with the seam bias. It is however not the focus of this dissertation to provide a full review of them. The interested reader is referred to a very comprehensive review by Lynn and colleagues (2005).

1.5 Seam effects in labor force transitions

It is useful to summarize the specific findings of seam effect in labor force transitions because those are the variables used in this dissertation.

In order to study labor market dynamics, individuals are coded into one of three mutually exclusive states for each month: employed (E), unemployed (U), and not in the labor force (N). Even if it is possible for an individual to be legitimately in two or three states in the same month\(^4\), many authors do not discuss this possibility. Some authors follow certain rules, for example, giving priority to E over U and to U over N (Cotton & Giles, 1998). The combination of states results in 9 pairs of classifiable codes, as delineated in Table 1.1.

Table 1.1
Possible Combinations of Mutually Exclusive States in Labor Force Status

<table>
<thead>
<tr>
<th>Month t</th>
<th>E</th>
<th>U</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>EE</td>
<td>EU</td>
<td>EN</td>
</tr>
<tr>
<td>U</td>
<td>UE</td>
<td>UU</td>
<td>UN</td>
</tr>
<tr>
<td>N</td>
<td>NE</td>
<td>NU</td>
<td>NN</td>
</tr>
</tbody>
</table>

After each subject is classified in one state for each month, status changes are computed for each of the nine combinations. In the LFS literature, the status changes on

\(^4\) It is the case, for example, of switching between employment and unemployment in the same month because the two events occurred at different weeks of the same month.
the diagonal of Table 1 (EE, UU, NN) are referred to as *stayers* (stay rate) or *non movers*. These people keep the same status from one month to the next. The remaining six status changes are referred to as *movers* (EU, EN, UE, UN, NE, NU).

The first application of seam effect analysis to labor force transitions was done by Martini (1989) using the Survey of Income and Program Participation (SIPP) data from 1984 and 1985. He noticed an increase in transition rates for the movers at the seam for every transition, as shown in Figure 1.8.

![SIPP 1984 labor force transition rates](image)

Figure 1.8 Labor Force Transition Rates for the SIPP. Data from Martini (1989, p. 390).

Seam effects with SIPP labor force transitions were subsequently studied by Martini and Ryscavage (1991) comparing the SIPP data to the CPS transition rates. Ryscavage (1993) compared seam bias in 1987 to 1990 finding an increase of the seam problem in the later waves. Lemaître (1992) showed an increase of three to four times in the num-
ber of transitions in and out of self-employment at the seam when compared to the rest of the years using the Canadian Labor Market Activity Survey (LMAS). All the six movers transitions in the LMAS were later studied by Cotton and Giles (1998) who found seam effects for most of the transitions. Torelli and Trivellato (1993a; , 1993b) showed strong seam (called heaping) effects for unemployment duration spells in the Italian Labor Force Survey. Heaping effects for inflow and outflow transitions are found by Kraus and Steiner (1998) in the German Socio-Economic Panel (GSOEP). Still in the European context, seam effect in transition rates are shown in the UK subsample of the European Community Household Panel (ECHP) by Jäckle and Lynn (2004).

1.6 Conclusions from the literature

In summary, seam effects are common across different types of variables and have been found in panels across the world, with data collected either by telephone or face-to-face, no matter the length of the reference period.

Seam effects are important because their magnitude prevents the use of the transitions at the seam as they are, but some statistical adjustment is necessary. From validation studies we learned that within wave transitions are underreported whereas between wave transitions are over reported.

Current explanation for the seam effect can be classified in four classes, data management issues, forms of satisficing, memory issues and respondent errors. These four potential causes may happen frequently and at the same time making difficult to isolate each of them in a regular panel unless special experimental studies are carried out.

The common denominator of the data collection methods that thus far have shown to be effective in reducing seam effects is that they use different memory aids in helping the respondent to retrieve information. Dependent interviewing, the use of a calendar, and using the same interviewer across waves all give respondents more retrieval cues than a standard panel setting. In the first case the respondent is reminded of the previous wave’s answer and/or inconsistencies are pointed out thus inviting the respondent to try harder in
retrieving the answers. The calendar visualizes the answers from the previous wave and locates them in time, working as a sort of diary and providing, in a different shape, retrieval cues similar to dependent interviewing. If the same interviewer administers the questionnaire in each wave, then the same interviewing setting is recreated, or, in other words, Tulving’s memory encoding specificity principle is used.

In addition, grouping the questions by reference period and not by item (Rips, Conrad, and Fricker, 2003) reduces satisficing behaviors and “forces” the respondent to think more about the events. Lastly, backward order of retrieval seems more effective even if previous research on recall order in survey has found mixed results.

In the next chapter the methodology of Event History Calendar (EHC) will be presented. EHC is a relatively new calendar aided data collection methodology (Belli & Callegaro, in press-a) that is being applied to very few panel designs. Currently PSID, HILDA, and the Italian Longitudinal Household Survey use EHC in some sections of the questionnaire. Findings from methodological studies measuring the quality of the data collected with EHC in comparison to CQ will be presented and discussed. The chapter will be concluded with a detailed description of the PSID computerized EHC instrument.
Chapter 2.

The Event History Calendar data collection methodology and its application in the PSID
2.1 A definition of the Event History Calendar

Calendar-based data collection originated in the late 1960s (Balán, Browning, Jelin, & Litzler, 1969; Blum, Karweit, & Sørensen, 1969) and was then rediscovered by Freedman and colleagues (1988) who called it the Life History Calendar5 (Belli & Callegaro, in press-a). Authors have used inconsistent terminologies to refer to the same interviewing strategy. For the purpose of this chapter I will not enter into the terminological details but refer to it as Event History Calendar or EHC. The interested reader can refer to Belli and Callegaro (in press-a) for more details.

An EHC interview is centered around a customized calendar that shows the reference period under investigation (Axinn, Pearce, & Ghimire, 1999). The calendar contains timelines for different domains, for example work history, residence history, household composition and other domains relevant to the topic of the study. Landmark events, such as holidays and birthdays can be used to aid the respondent’s memory (Belli, Shay, & Stafford, 2001) but not every author employs them (Axinn, Barber, & Ghimire, 1997). The interviewer guides the respondent in answering queries for each time line, starting with the first domain and continuing down until the last domains, which are the focal points of the study, are completed. The EHC also allows moving back and forth among the domains. The process uses previous domain information and dates to help the respondent correctly place other events in the appropriate time frame. When, for example, unemployment history is queried, the respondent can use retrieval cues from landmark events, residential history, and household composition to retrieve the spell in which any unemployment had transpired. For instance, an unemployment spell can happen just before a move to a new location, or right after a pregnancy. The calendar does not necessarily have to be shown to the respondents. In the literature there are successful applications using both face-to-face (Axinn, Barber, & Ghimire, 1997; Parrado & Zenteno, 2001) and telephone interviews (Belli, Shay, & Stafford, 2001; Belli, Smith, Andreski, & Agrawal, 2006). In a randomized assignment study, Callegaro and colleagues (2005) found initial

5 The work of Freedman et al. (1988) is the most cited paper in the EHC literature.
support that showing the calendar improves the quality of reports in comparison to a completely auditory presentation (checked against validation data). In a computerized self-administered panel, respondents were given a summary screen in the form of a calendar based on previous responses collected with CQ about major events in various domains of their life. After the summarized calendar screen the respondents were asked to make any necessary correction. Twenty one percent of respondents corrected the order of the events and seventy nine percent corrected the dates after seeing the information summarized on the calendar (Saris, 1998).

The units of analysis are decided by the researcher and have included years (Axinn, Barber, & Ghimire, 1997; Furstenberg, Brooks-Gunn, & Morgan, 1987), months (Caspi et al., 1996; Engel, Keifer, & Zahm, 2001), and thirds of a month (Belli, Shay, & Stafford, 2001; Pebley, 2004a). In some administrations, interviewers are encouraged to permit a chronological order of retrieval at the respondents’ choosing such as moving forward in time, backward in time, or across contemporary events. Every event or “spell” is entered into the time lines with a starting point and an ending point as defined by the unit of analysis. In case the respondent remembers the exact dates those can be entered into the timeline or are converted directly into the unit of analysis. If those are not available, the interviewer is instructed to probe for a time location depending on the unit of analysis. In the case of thirds of a month, for example, the interviewer will ask if the event occurred at the beginning of the month, in the middle, or at the end of that month.

Each spell can happen before, after, or contemporaneously with another spell. For example, a relocation event can occur contemporaneously with a new job event. The entire process of compiling the calendar focuses, by its nature, on coherence, consistency, sequential order, and attempts to correct for missing data (Belli, 1998). By coherence, it is intended that some events are less likely to occur together, such as having three jobs at the same time. Although that might be the case, it is also possible that the respondent made a mistake in the location of the jobs in time. The calendar instrument visualizes better than a CQ instrument events in the time line suggesting possible inconsistencies. The sequential nature of the EHC is revealed by the fact that an event should happen after the
preceding one and before the following one. For example if the respondents are unemployed, they will then look for a job, and when found, become employed. The time line also highlights to the interviewer missing data in a more prominent way than a CQ. This is because each event in a time line should be adjacent to another event without any time unaccounted for (something must have been happened in each time frame). In case of time unaccounted for, the interviewer can probe to investigate what happened in that time frame. Figure 2.1 shows the PSID paper and pencil calendar method study instrument (Belli, Shay, & Stafford, 2001).
Figure 2.1 Reproduction of the Instrument Used in the 1997 PSID Paper and Pencil Calendar Method Study.
As we can see, the instrument is organized by domains (rows) and by months (columns). This calendar collects information for a two year reference period and the units of analysis are thirds of a month. It starts by asking for some landmarks events, some of them are already filled in such as national holidays (e.g. Labor Day and Thanksgiving). The left columns leave plenty of space to enter information that otherwise won’t fit in the timeline, they work as a sort of schematic questionnaire. For example in the employment domain of Figure 1, the left column is used to record the description of each job and the name of the employer and in the last domain (welfare receipts) there are yes/no questions that are answered with a checkbox. The right side is devoted to deal with start and end of spells. The interviewers are instructed to use a special notation to fill out the timelines, for example an X is the start and end of a spell and in between dashes ( -- ) are used. In designing this PSID calendar, Belli et al. were informed by the work of Freedman and colleagues (1988).

The calendar is accompanied by a questionnaire that provides a structure, suggests question wording and probes. The questions can be very detailed and standardized (Zahm et al., 2001) or more script-like and open ended (Blum, Karweit, & Sørensen, 1969). The common denominator of EHC studies is that the interviewers are not required to follow the question order exactly, but they are allowed to deviate from it with a series of probes and inquiring techniques that are adapted to the respondent’s way of remembering. The instrument’s flexibility promotes the recording of very complicated life histories, with the benefit of allowing the respondents to remember the events in the chronological order which they preferred. Although the literature about which order of retrieval (forward or backward in time) leads to the best reports is still inconclusive (Bhandari & Wagner, 2006), some studies on the quality of order of retrieval suggest that when allowing respondents to choose their own order in comparison of suggesting an order of retrieval provides better data (Jobe et al., 1990; Loftus & Fathi, 1985). The EHC instrument does not necessarily have to be used by itself. In fact, it is frequent that it is part of a longer instrument in which a CQ component is administered before and/or after the calendar (Belli, 2003; Pebley, 2004b; Witterbrood & Nieuwbeerta, 2000).
2.2 Quality of EHC data

Since its first applications, researchers have been examining the quality of data obtained by EHC interviewing and how it addresses the common problems of recall error in survey research (Belli, 1998). The quality of the EHC has been assessed by matching EHC with validation data, via test-retest of inter-respondent reliability methods, and by comparing EHC to CQ data of respondents randomly assigned to those conditions.

Freedman and colleagues (1988) published the first paper dealing with quality of EHC data by comparing EHC reports with validation data. These data were contemporary verbal reports collected from the same panel respondents five years earlier. For events such as marital status and births the agreement was above 95% with discrepancies of one month. For school attendance and employment status measured on a 3-category scale (full-time, part-time, nonattendance/unemployment) the reliability was of .90 and .62 respectively (Alwin, 2007). Caspi et al. (1996) found over 90% agreement of data reported in the same month three years earlier for living arrangements, cohabitation, schooling, employment, and job training in a longitudinal-epidemiological study. Similar rates of agreement with data collected 15 years earlier were found by Ensel and colleagues (1996) in a longitudinal study. Lin, Ensel & Lai (1997) asked 1993-95 Albany panel respondents to recall life experiences that happened in 1979-82 using Event History Calendar data collection. They then compared the EHC reports with the on-time reports of the same respondents for CQ data collected nearly 15 years before. The comparisons were done for ten life event categories such as marital changes, births, residential moves, education, employment, deaths, financial/legal work, health, sexual difficulties, arguments with partner or spouse, and changes in social activities and recreation. The analysis showed nearly no overreporting but some underreporting. The underreporting errors were no higher than 33% for health events and tended to be larger for chronic or routine activities rather than intimate events and family experiences. Interestingly, 25% of respondents made no recall errors in any of the ten categories and another quarter made recall errors in only one of the ten categories. Only 8% made recall errors in five or more categories.
When analyzing the precision of recall over time, the authors found nearly no backward or forward telescoping, concluding that recall errors were mostly due to forgetting or selective retention, rather than timing of their occurrence.

In a test-retest study of an icon-based EHC among farm workers, the subjects were interviewed twice, 8-14 months apart, about their lifetime employment. The icons represented a pictorial representation of the subject’s work/life and proved particularly effective for illiterate or partially-illiterate subjects. The inter-calendar agreement was moderate to high across all periods for certain crops, for location, and for agricultural work in general. Agreement for job counts was high for certain crops, e.g. .93 for apple-related work (Engel, Keifer, Thompson, & Zahm, 2001). Figure 2.2 shows the calendar used in the study.

![Calendar used in the Engel et al. (2001) Study of Migrant Farm Workers, in the public domain.](image)
Kessler and Wethington (1991) compared inter-respondent reliability of data collected with the EHC with a similar study using a CQ. The respondents were married couples who were separately asked to date severe negative events such as illness, death, and job and financial troubles, for the previous 12 months before the interview. The interreliability rate \( r \) among the couple was in the range of .64 to .68 compared to .30 for a similar CQ study.

When compared to CQ methods in randomized design studies, EHCs have shown better data quality for retrospective reports in terms of precision of the placement of events in time, and in terms of reducing underreporting. In a study with PSID respondents randomly assigned to EHC or CQ conditions, and using validation data collected from the same respondents (PSID) years before, Belli, Shay, and Shafford (2001) showed that EHC reports were more precise (less underreporting, higher agreement between reports) than CQ reports on number of moves, income, weeks unemployed, and weeks missing work resulting from personal illness, the illness of another, or the combination of the two. In a study with a design similar to that of Belli and colleagues (2001) with PSID respondents, this time using a computerized instrument and telephone interviews, Belli, Smith, Andreski and Agrawal (2006) collected retrospective reports with a reference period of up to 30 years. The respondents were randomly assigned to either a standardized CATI interview or a computerized EHC (C-EHC). The reports on social and economic variables of residential, marriage, cohabitation, and work history were compared to the previous data collected from the same panel respondents. The C-EHC showed better overall data quality for cohabitation and work history; no difference was found for residence change and CQ showed better data only for marriage history. Yoshihama and colleagues (2005) found that the EHC method facilitated the recall of domestic violence victimization. In particular, the EHC technique elicited more reports of lifetime experiences of intimate partner abuse (93-96%) than CQ format (61-64%). EHC was especially precise in revealing abuse that occurred earlier in respondents’ lives.
2.3 EHC and the structure of autobiographical memory

EHC enhances the quality and the precision of retrospective reports because it taps three basic types of mechanisms that are associated with the structure of autobiographical memory: “top-down”, sequential and parallel cueing (Belli, 1998). According to Barsalou (1988) and Conway (1996), autobiographical memory is organized hierarchically where at the top of the hierarchy exist memories of lifetime periods, arranged in thematic domains (family, career, education, etc.), and which are connected to the self-concept of the person. Temporal changes are stored for each domain, for example, the transition from student to worker within the career domain. In the middle of the hierarchy reside memories for general events, and at the bottom reside memories for specific events (see also Conway and Rubin (1993) for an extended description). The EHC calendar is generally organized in such a way that it moves from the general to the specific (top down), which matches the natural organization of autobiographical memory. At the same time, respondents are provided a sort of roadmap with which to organize his/her memories. Moreover, the structure of the EHC calendar encourages the respondent to fill in temporal gaps and to explain multiple events which happened at the same time.

Top down cueing taps into the relationship between the top of the hierarchy and its bottom, where, for example, general events are nested into a self-concept lifetime period. Sequential cueing taps into the chronological sequence of events that belong to the same domain. Parallel cueing taps into the links that exist across domains and themes. The flexible interviewing style of the EHC method allows for the proper memory cues and adapts itself to the current interview. The respondent feels more engaged in the conversation because each question is “personalized” or tailored. Furthermore, the entire interview is conducted in a narrative style, more compatible with the manner that autobiographical memory is structured and retrieved (N. R. Brown & Schopflocher, 1998; Schank & Abelson, 1995). Flexibility, adaptability and tailoring enhance comprehension (Conrad & Schober, 2000; Schober & Conrad, 1997, 2002) by allowing the full
resources of communication in a collaborative construction of meaning (Suchman & Jordan, 1990).

There is also another byproduct of the EHC technique that enhances the quality of reports. During an EHC interview, it may be more difficult to satisfice (Callegaro, Yu, Cheng, Liao, & Belli, 2005). When faced with CQ survey questions, people can use a satisficing approach (Krosnick, 1991) in order to provide a plausible answer. Satisficing respondents do not try to fully understand the question, but rather understand just enough to form a plausible answer. They also do not fully retrieve all the available information from memory but often only the minimum necessary on which to base an answer. The time line of the calendar asks for coherence and precision, discouraging the guessing and the reporting of merely plausible answers. It also encourages that the reported information is exhaustive, and, in its computerized version, automatically highlights inconsistencies to give feedback to the respondent about which sections still require more information (Belli, 2000). Moreover, the narrative style and the tailoring of the EHC method enhance the motivation, engagement, and interest of the respondent, other factors responsible for optimizing strategies (Krosnick, Narayan, & Smith, 1996).
2.4 The PSID computerized EHC

When the PSID moved to EHC interviewing in 2003, a computer-assisted interviewing EHC instrument, copyright to the Regents of the University of Michigan, was developed and integrated with the computerized Blaise® CQ instrument (Belli, 2003). The interviews started in CQ mode, switched to C-EHC and then returned to CQ as shown in Figure 2.3.

Figure 2.3 Schematic PSID Questionnaire

The C-EHC was designed to collect information in five domains, Landmark Events, Residence, Employment, Not Working, and Time Away. Figure 2.4 depicts a screen capture of the PSID 2005 C-EHC instrument.
The reference period is two years, starting, in this case, from January 2003 to December 2004. The calendar also depicts seasons and months (two rows) to be used as temporal cues in facilitate remembering whenever the respondent is uncertain about the timing of events. The top section provides a summary of key events that have been already recorded in earlier parts of the interview, to facilitate parallel retrieval. It is updated continuously but does not change in format. To highlight available specific retrieval cues, the interviewer places the cursor on a specific calendar node. A pop up message will appear with the requested information, for example where the respondent was living at that time. The middle section is used to enter data within each domain. Reflecting the structure of autobiographical memory, this section consists of hierarchically organized timeline domains. Each domain is represented by a “tab” organized in a left to right order (Belli, 2000). Each tab brings up each of the domains to be covered: landmarks, residence, employment, not working, and time away. The bottom section is used for addi-
tional data entry. It contains tabs with introductory scripts for each domain. It displays a data entry matrix that interacts with the timeline where the interviewer can enter information into the timeline of the middle section, for example the name of the employer. Figure 2.5 shows the residential tab domain ready to accept data entry.

Figure 2.5 Residential Domain Screenshot

When a domain is currently used it is displayed in bold. In this case the software shows the screen when the interviewer is entering the start data and the end date of a residential spell. The current button is used to automatically enter the current date for the ending time. The bottom part of the screen displays residential questions in a CQ style to be entered before moving to the next residential spell.

Potential inconsistencies and gaps are automatically flagged by the software using a red or white color. In the first case, when there is an overlap of two time lines, it is
flagged in red. One example can be of a respondent indicating being working and not working for the same period of time. Interviewers are trained to spot gaps in every domain (flagged with white blank spaces). In that case they are instructed to probe and trying to fill the gaps. An example can be that the respondent reported the wrong dates of a move and correcting that will close the gap.

2.4.1 *EHC techniques*

Interviewers are instructed to use a number of techniques when entering the data into the C-EHC instruments. Four of them are particularly important: the narrowing technique, the how-long-ago verification technique, the parallel probing, and the freedom in chronological order of retrieval (PSID staff, 2005b).

Because the units of analysis are thirds of a month, interviewers are instructed to probe appropriately in case a respondent does not remember the exact date or remembers a week that can be classified either in the first or in the second third of a month (e.g. the second week of February). In this case the standard probe is: Would you say that it was more toward the beginning, the middle, or the last part of the month?” The narrowing technique is connected the concept of thirds of a month as unit of analysis. It is used when the respondent is unsure when an event occurred. The procedure is to first ask about what year and season or quarter the event occurred and then to move a month level. Finally the event is located into one of the thirds of a month. As an example consider the following: “You said you moved to this home in the spring of 2003. Would you say that was in March, April, or May?” Then: “You said you moved in April. Was that the beginning, middle or end of April?”

The *how-long-ago verification* technique is used to probe or verify dates for respondents that may confuse years when reporting events. For example, if the interview is in February 2005 and the respondent reports that he changed jobs in November 2003, the interviewer should ask, “So that happened in November 2003, just over a year ago. Is that correct?” Sometimes the respondent will say “yes”. Other times, they may say they actu-
ally meant November 2004, only a few months before the interview. The how-long-ago technique is aided by the software; if the interviewer places the cursor on any month on the calendar, a pop-up box will show how long ago that month occurred.

Interviewers are also trained to use *parallel probing* techniques while administering the instrument. Landmarks, events, residence changes, and holidays are very useful to cross check the dates of other events. For example a probe can be: “I see that you became unemployed and then you moved a couple of months afterwards, is that correct?”

The *freedom in chronological order of retrieval* technique means that respondents are allowed to report events in any order they want. For example interviewers are instructed to let respondents report landmark events in any order they think of them, instead of forcing for a particular backward or forward in time order. For other domains the majorities of respondents remember events and dates with recent events and then move backward in time. But if some respondent reports events in a different order, interviewers should allow them doing so, paying attention that all events are in fact reported and that there are no gaps in the timeline.

In the next chapter I will describe the sample design, the eligibility and following rules of the Panel Study of Income Dynamics. This will serve to get the reader familiar with the terminology used in the PSID and later on in the dissertation. Data collection practices, response rates, and interviewing rules will also be presented for their impact on the data analyzed.
Chapter 3.

The Panel Study of Income Dynamics (PSID) data collection procedures
3.1 The PSID: the first household panel in history

In this chapter the reader will become familiar with the PSID data collection procedures and the changes that have been implemented during the life of the panel. The chapter will be useful to better understand the terminology and the results of the analysis presented in the next section of the dissertation.

The first wave of the PSID was administered in 1968 on a representative sample of U.S. individuals and the family units in which they reside. The PSID is the first household panel in history (Ruspini, 2002). The following description of panel design, sampling, following rules, data collection, response rates, field and interviewing procedures is compiled from different sources (Beckett, Gould, Lillard, & Welch, 1988; C. Brown, Duncan, & Stafford, 1996; Duncan, 1999; M. S. Hill, 1992; McGonagle & Schoeni, 2006; PSID staff, 2003, 2005a, 2005b, 2006a, 2006b).

3.1.1 Sample design

The original 1968 PSID sample was composed of two independent samples: a cross sectional national sample of approximately 3,000 households, and a national sample of low-income families of approximately 2,000 households. The first sample was designed by the Survey Research Center of the University of Michigan (the “SRC sample”) and was an equal probability sample of households in the 48 contiguous states, designed to yield approximately 3000 completed interviews (precisely 2,930 in the first wave of 1968). The second sample was selected from the 30,000 families of the Survey of Economic Opportunity (the “SEO sample”) conducted by the Bureau of Census for the Office of Economic Opportunity. The PSID selected about 2,000 low-income families with heads under the age of 60; precisely 1,872 families were successfully interviewed. The inclusion criterion was that the household had an income at or below 150 percent of the pov-

---

6 In order to facilitate the reading of this chapter I decided to put the majority of references about the PSID design at the beginning of the chapter, instead of continuously inserting them in the text.
The SEO was confined to Standard Metropolitan Areas (SMSA’s) in the North and non SMSA’s in the Southern region. The SEO sample resulted in a sizable subsample of African Americans.

In 1990 the PSID added 2,000 Latino households, including families originally from Mexico, Puerto Rico, and Cuba. The Latino sample was however dropped in 1995 because of lack of funding and also because it missed a full range of post-1968 immigrants, Asians in particular.

In 1997, in order to keep the study representative of the U.S. population, and also in order to accommodate the study five-year funding, a refresher sample of 441 post 1968 immigrant families and their adult children was introduced. Some of these families did not meet eligibility criteria to be interviewed in 1997 but became eligible in 1999, reaching the number of 511. In 1997 the core sample was also reduced. More specifically the core sample was reduced from approximately 8,500 families in 1995 to 6,168 in 1997. The majority of cuts were taken from the SEO sample, of which, 43% remained in the active sample. Immigrants for the refresher sample were eligible if they had immigrated to the United States after 1968 and who were not married to persons who were living in the U.S. at the time of the original 1968 PSID sample selection.

When using the weights designed to compensate for both unequal selection probabilities and differential attrition, the PSID remains representative of the U.S. population for a given temporal snapshot.

3.1.2 Definitions of eligible sample members

The PSID collects information about family units (FU). The FU is defined as a group of people living together as a family. They are generally related by blood, marriage, or adoption, but unrelated persons can be part of a FU if they are permanently living together and share both income and expenses. Families change from year to year. Each FU has one and only one Head. In a married-couple family the head is the husband, unless the husband is severely disabled. This definition was adopted from the Bureau of Census
in 1968 and, although dated, is has been maintained to keep consistency and ease of following panel members. The person designated as head can change over time. When a new head has to be chosen the following rules apply: the head of the FU must be at least 16 years old and the person with the most financial responsibility for the FU. If this person is female and she has a husband in the FU, then he is designated as head. If she has a boyfriend with whom she has been living for at least one year, then he is head. However, if the husband or boyfriend is incapacitated and unable to fulfill the functions of head, then the FU will have a female head.

A new head is selected if any of the following conditions apply: (a) last year's head moved out of the HU (household unit), died or became incapacitated; or (b) a female head has gotten married; or (c) if this is a new FU. The person living with the head is defined as Wife if legally married or “Wife” if she is cohabitant. A cohabitant is labeled a boyfriend or girlfriend the first wave he or she appears in the study. If that cohabitant is still in that same family unit at the time of the subsequent interview, the cohabitant's label switches to "wife" if the cohabitant is female; if the circumstances are otherwise the same and the cohabitant is male, his label switches to head and his female partner (who has been head) becomes a "wife." In any case the boyfriend becomes head and the girlfriend becomes “wife” only if they have been living in the FU for at least one year. This is called the “one year rule” and is one of the most important PSID family composition rules. Boyfriends and girlfriends are treated like family members who are not Heads or Wives/"Wives," and some information is obtained about them. Individuals are classified as family members if, at the time of the interview, they are either living in an interviewing FU (and not temporarily roommates, guest, visiting friends and relatives) or temporarily away from the FU because they are living in an institution.

3.1.3 Following rules

Following rules are designed to follow up and update the initial sample and dictate who is eligible of being interviewed. The goal is to make sure that the sample remains cross-
sectionally representative of the population. Without following rules the initial sample will age up to the point where no respondent will be available anymore.

The PSID follows members of the original family units and their adult offspring. In other words, children born to sample members after 1968 inherit the “PSID gene” and are followed as well. PSID distinguishes between sample person and nonsample person. A sample person is an individual belonging to the original samples or eligible to be interviewed in accordance to the following rules (e.g. splitoff). Nonsample persons entered the sample after 1968 through marriage or a living arrangement that places them in the same household. Nonsample person are not followed when they stop residing with a sample person. PSID also assign a weight of 0 to nonsample person.

A family member who moves out of the original PSID family is eligible of being interviewed if he or she is an adult sample member and living in an independent household. These members form the so called splitoff families. A splitoff family consists of a person or group of people (at least one of whom is a "follow" person of any age) who moved out from a main family since the prior wave's interview to form a new, economically independent family unit. Several criteria must be met for a splitoff to occur. In addition to having moved out since the prior wave, and to being 'followable' (i.e., having an original 1968 family identification), the person or group of people in general may not have moved to an institution such as college or prison or to another family unit within the panel study. Moreover, the person or group of people who have moved out and formed their own family unit must be economically independent (i.e., they must be paying their own living expenses) from the family unit from which they split off. Children born to original sample members are eligible to be followed as separate FU when they move out of the original PSID FU. Ex-spouses and other adult sample member who move out of the PSID FU are followed to their new family units.

In some cases PSID interviews members under the age of 18 who have set up their own household. If a family members 18 years and older moves to an institution (college, military, prison) they are not interviewed, but PSID keeps track of them (tracing rules). Interviews with those members are attempted if and when they leave the institu-
tion to set up their own households. Table 3.1 provides some examples of the PSID following rules for special cases.

Table 3.1
PSID Following Rules Examples for Special Cases

<table>
<thead>
<tr>
<th>Situation</th>
<th>Followable?</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 16 year old sample-member daughter of a FU, gets married and lives in</td>
<td>Yes</td>
</tr>
<tr>
<td>a household other than her parents</td>
<td></td>
</tr>
<tr>
<td>A 16 year old sample-member son moves out of a FU but is still</td>
<td>If he has not entered an institution,</td>
</tr>
<tr>
<td>economically supported by his parents</td>
<td>he will be followed to his new</td>
</tr>
<tr>
<td></td>
<td>household unless he has moved in with</td>
</tr>
<tr>
<td></td>
<td>a currently response PSID family. If</td>
</tr>
<tr>
<td></td>
<td>the latter, he will simply be</td>
</tr>
<tr>
<td></td>
<td>incorporated into the new FU. However,</td>
</tr>
<tr>
<td></td>
<td>in either circumstance, it is unlikely</td>
</tr>
<tr>
<td></td>
<td>that he will be treated as head.</td>
</tr>
<tr>
<td>A sample member goes to college</td>
<td>No, he/she will be interviewed when</td>
</tr>
<tr>
<td></td>
<td>out of college and forming a new</td>
</tr>
<tr>
<td></td>
<td>economic independent household; if</td>
</tr>
<tr>
<td></td>
<td>he/she returns to the parental family</td>
</tr>
<tr>
<td></td>
<td>or goes to live with another family</td>
</tr>
<tr>
<td></td>
<td>member who is also being interviewed,</td>
</tr>
<tr>
<td></td>
<td>he/she will simply be incorporated as</td>
</tr>
<tr>
<td></td>
<td>a mover-in to that family.</td>
</tr>
<tr>
<td>A sample member goes to jail and then returns back to the household</td>
<td>PSID keeps track of him/her when in</td>
</tr>
<tr>
<td></td>
<td>jail and incorporates him/her into the</td>
</tr>
<tr>
<td></td>
<td>main family unit when he/she is released.</td>
</tr>
<tr>
<td>A FU gets separated and the husband and wife form new economically</td>
<td>Yes, each is followed separately if a</td>
</tr>
<tr>
<td>independent households</td>
<td>sample member or otherwise followable</td>
</tr>
<tr>
<td></td>
<td>(e.g., non-sample parents of sample</td>
</tr>
<tr>
<td></td>
<td>children under age 25 were followed</td>
</tr>
<tr>
<td></td>
<td>1994-2003). Children of the previous</td>
</tr>
<tr>
<td></td>
<td>FU are followed as well if they are</td>
</tr>
<tr>
<td></td>
<td>sample members.</td>
</tr>
<tr>
<td>A sample person divorces and gets remarried. His second wife moves in</td>
<td>She is almost certainly non-sample. If</td>
</tr>
<tr>
<td></td>
<td>she moves out, she is not followed</td>
</tr>
<tr>
<td></td>
<td>(unless she meets one of the various</td>
</tr>
<tr>
<td></td>
<td>rules over the years for following</td>
</tr>
<tr>
<td></td>
<td>non-sample persons—see above).</td>
</tr>
</tbody>
</table>

---

7 I thank Tecla Loup for her assistance in compiling Table 3.1.
3.2 Data collection

Data were collected face-to-face with paper and pencil questionnaires until 1972. In 1973 data started being collected over the telephone, still with paper and pencil questionnaires and with interviewers conducting interviews from their own homes. In person interviews have been collected only with respondents who do not have a phone or where telephone interviewing is problematic, roughly 500 families from the 1973 to 1992 wave. Starting in 1993, data were collected using computer assisted telephone interviewing from a centralized location, specifically the call center of the Survey Research Center of the University of Michigan. In the 1999 wave, 97.5% of the data were collected over the phone and all interviews were conducted using computer based instruments. During the last waves, telephone data collection has been decentralized, and nowadays nearly all sample members are interviewed by phone by interviewers in the field, though there are a few cases that require a personal visit.

The PSID collected data annually until 1997, at which time the panel moved to a biennial data collection. Beginning in 2003, a computer-assisted Event History Calendar (C-EHC) instrument was integrated with the current CATI instrument (Blaise). Five sections of the questionnaire (domains) were collected with C-EHC, particularly landmark events, residential, employment, not working, and time away histories. The new instrument was pretested in 2002 where some panel respondents were randomly selected and 100 interviews were conducted. In the 2005 the same five sections of the 2003 questionnaire were collected with C-EHC. Figure 3.1 summarizes the major changes in design of the PSID from 1968 to 2005.
Figure 3.1 PSID Major Design Changes Timeline
3.2.1 Response rates

In 1968, 76 percent of the households from the sample were successfully interviewed, and 86 percent of the 76 percent were interviewed in the second wave. Since then the PSID response rate kept being very high with an average core reinterview rate of 96-98%. Nevertheless by 1989, the PSID had experienced approximately 50 percent loss from cumulative attrition. The effect of cumulative attrition has been studied by Fitzgerald, Gottschalk & Moffitt (1998). The response rate of the most recent waves has been high, despite the rapid decline of most cross sectional surveys in the U.S. Table 3.2 reports the response rate for the last two waves.

Table 3.2
Response Rates and Sample Sizes for the 2003-2005 Waves

<table>
<thead>
<tr>
<th>Sample</th>
<th>Non Splitoffs</th>
<th></th>
<th>Splitoffs</th>
<th></th>
<th>Total # of interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Response rate (%)</td>
<td>N of interviews</td>
<td>Response rate (%)</td>
<td>N of interviews</td>
<td></td>
</tr>
<tr>
<td>2003 Actual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core re-interview</td>
<td>97</td>
<td>6554</td>
<td>83</td>
<td>561</td>
<td>7115</td>
</tr>
<tr>
<td>Core re-contact</td>
<td>65</td>
<td>200</td>
<td>56</td>
<td>15</td>
<td>215</td>
</tr>
<tr>
<td>Core subtotal</td>
<td>95</td>
<td>6754</td>
<td>82</td>
<td>576</td>
<td>7330</td>
</tr>
<tr>
<td>Immigrant re-interview</td>
<td>94</td>
<td>459</td>
<td>61</td>
<td>36</td>
<td>495</td>
</tr>
<tr>
<td>Immigrant re-contact</td>
<td>51</td>
<td>42</td>
<td>43</td>
<td>3</td>
<td>45</td>
</tr>
<tr>
<td>Immigrant subtotal</td>
<td>88</td>
<td>501</td>
<td>59</td>
<td>39</td>
<td>540</td>
</tr>
<tr>
<td>Total</td>
<td>95</td>
<td>7255</td>
<td>80</td>
<td>615</td>
<td>7870</td>
</tr>
<tr>
<td>2005 Actual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core re-interview</td>
<td>98</td>
<td>6756</td>
<td>89</td>
<td>516</td>
<td>7272</td>
</tr>
<tr>
<td>Core re-contact</td>
<td>61</td>
<td>194</td>
<td>60</td>
<td>3</td>
<td>197</td>
</tr>
<tr>
<td>Core subtotal</td>
<td>96</td>
<td>6950</td>
<td>88</td>
<td>519</td>
<td>7469</td>
</tr>
<tr>
<td>Immigrant re-interview</td>
<td>93</td>
<td>500</td>
<td>74</td>
<td>45</td>
<td>545</td>
</tr>
<tr>
<td>Immigrant re-contact</td>
<td>42</td>
<td>27</td>
<td>-</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>Immigrant subtotal</td>
<td>88</td>
<td>527</td>
<td>74</td>
<td>45</td>
<td>572</td>
</tr>
<tr>
<td>Total</td>
<td>95</td>
<td>7477</td>
<td>87</td>
<td>564</td>
<td>8041</td>
</tr>
</tbody>
</table>

Since the first wave, respondents were given a monetary incentive as a token of appreciation for the interview. The incentive for answering the questionnaire started at $5 in the first wave and reached $60 in the 2005 wave. A booklet explaining how the data have been analyzed and used is sent to PSID FUs. Other material such as announcements, respondents’ reports, and announcements of the upcoming interview is also mailed first class to each FU. For refusals, or persons reluctant to participate in future waves, a personalized persuasion letter is sent in order to increase response rates and decrease panel attrition. Beginning in the 1992 there have been routine attempts to reinterview prior-wave nonrespondents. These efforts were expanded in 1994 when a major effort was launched to bring back nonrespondents into the study, some of them who had not been interviewed since 1968. Nowadays, at each wave, there are attempts to interview previous wave nonrespondents. These cases are called recontacts. The PSID also tries to keep the same interviewer throughout the waves for the same FU. Figure 3.2 presents a summary of the flows of people that can be in and out of the panel.
Respondents can enter the panel from three sources: the original 1968 sample \((N)\); the 1997 refresher sample of post-1968 immigrants \((n)\); and births and marriages in existing families \((d)\). Children who split off as adults to form their own households \((s)\) constitute the splitoff families and provide the intergenerational element to the panel. Respondents can exit the panel for two reasons, because they attrite \((a1, a2)\) or because of mortality \((m1, m2, m3, m4)\). By attempting to re-contact \((r1)\) families refusing to be interviewed or not located \((A1\) and \(A2)\) in previous waves, the active panel can be boosted.
3.2.2 Panel maintenance, tracing and tracking rules

Following Kalton and Citro (1993) and Trivellato (1999) definitions, the term *tracking* is used for procedures set up to keep track of the movements of panel members, and *tracing* methods are referred to strategies set up to find lost panel members.

PSID uses two methods to track respondents, an updated address and contact person is recorded at each interview, and an incentive for sending a keep-in-touch postcard with the updated address plus the name of the contact person is sent to panel members. The incentive for the keep-in-touch postcard started at $5 and was of $10 in the last 2005 wave. At the end of each interview, names and telephone of friends and relatives who would know the respondent’s location and contact information are recorded in an apposite section of the questionnaire.

Since 1993, tracking and tracing data are managed by internally developed software called SurveyTrak that administer the sample and administrative information about the FU. For the 2005 and successive waves, an Oracle-based respondent address payment system (RAPS) manages multiple respondents addresses, secondary addresses, address to which sent the incentive and generates tailored persuasion letters and all the personalized correspondence between the PSID and the panel members. In 2005 most respondents received the incentive payment within 48 hours.

3.2.3 Field procedures and interviewing rules

The PSID data are collected by the interviewing staff of the Survey Research Center of the University of Michigan. The field period generally starts in March and continues until September, October, or November, with some yearly variations and exceptions, for example, the 1993 and 94 wave interview period extended until December. The extended data collection period gives considerable time for interviewers and PSID staff to locate respondents that cannot be found on first contact.
Information is gathered about all persons residing in the FU but there is only one respondent per household. Most detail is collected for the head, substantial detail is collected for the wife/"wife" and some other detail is collected for the other family members.

Interviewers used to be instructed to prefer interviewing the head, whenever possible. The second choice in answering the questions would the wife or “wife”. This is no longer the case. Now the interviewers are trained to follow rules such as interviewing who is knowledgeable about the family finances and composition, who is willing to speak to them, and finally, who is available. Interviewers are also instructed to make their best attempt to interview the respondents from the last wave. Increasingly the PSID is finding that couples in a FU “switch off” being the respondent – one year the usual respondent might be too busy and the other half of the couple is available and willing. The use of interviews other than heads and wives is discouraged and interviewers are instructed to contact the survey manager before doing do. Examples of other respondents are grown children or other relatives, social workers, attorneys, or others who know enough about the financial situation of the FU.

3.2.4  Interviewer training and monitoring

Interviewers are extensively trained before each wave. They receive some materials, most importantly the Study Guide and the Question by Question Objectives (Q by Qs). The Study Guide introduces and defines all the PSID study concepts and conventions, the sample management tools, the questionnaire and all the field procedures. The Question by Question objectives manual describes each section of the questionnaire with definitions and examples for specific questions and sequences. This is an important feature of the PSID and makes it different from other studies. In fact PSID encourages directive probing when the respondents have comprehension problem. The following is an excerpt from the 2003 Study Guide (p. 6):
“We have specific ideas behind our questions. All respondents need to have the same understanding of terms used throughout the questionnaire. If a respondent is unsure how to answer a particular question, please do not say “whatever it means to you”. Instead, refer to the definitions provided in the Q by Qs. (There are a handful of exceptions to this rule: the “whatever it means to you” probe is allowable in a very few circumstances, which are clearly noted in the Q by Qs.)”

The above statement is a clear example of allowing the interviewers to use a more flexible interviewing style than a traditional standardized interview. In fact the PSID makes the point that the panel interview is different, and that the requirements of the study may run contrary to the training given by the Institute of Social Research for other projects.

Part of the Study Guide (9 pages) is devoted to train interviewers in handling refusals and in refusal conversion techniques. The section treats topics such as some general suggestions, provides some useful opening sentences, offers answers to common refusal sentences such as “I am too busy and don’t have time” or more specific ones such as: ‘last wave’s interview was too long”. The chapter also instructs how to deal with answering machines.

Interviewers are required to conduct some practice interviews with their team leader. Each interviewer will have 10% of their interviews verified and he/she is required to tape one every 10th interview. A portion of respondents are recontacted for monitoring and quality control. Interviewers’ supervisors recontact 5% of completed interviews for experienced interviewer and 10% for new interviewers. The verification consists in asking respondents 10 key questions such as how well the interview was conducted, how long it took, and some other questions from the original questionnaire.

The next chapter will start with a description of the hypotheses of this dissertation based on the studies previously reviewed. The dataset will be then presented, starting from a description of the operationalized variables used to compute labor force transition
rates. Primary findings on seam effect are discussed and interpreted. Secondary findings analyzing predictor variables of seam effect will conclude chapter 4.
Chapter 4.

Labor force transitions and seam effects in the PSID
4. 1 Study’s hypotheses and data analysis plan

4.1.1 Main hypothesis

The general assumption without which the main dissertation hypothesis cannot be formulated is the following: seam effect for labor force status (LFS) changes will be found in the PSID. To the author’s knowledge there is no previous study using PSID data documenting seam effect for labor force status changes. It is expected to find seam effects for two main reasons. First, because the PSID does not implement any strategy to reduce possible seam bias. Second, because seam effects for LFS has been found in similar panels with comparable question wording in the U.S. such as the SIPP (Fitzgerald, 1998; Martini, 1989; Nielsen & Gottschalck, 2006) and outside the U.S such as the SLID (Lemaître, 1992), the BHPS (Jäckle & Lynn, 2004), the GSOEP, (Hujer & Schneider, 1989), the HILDA (Carroll, 2006) and the Italian LFS (Torelli & Trivellato, 1993b). The main conclusion coming from all the literature on seam bias is that no panel is exempt from it although its magnitude depends on the variables studied.

Previous studies comparing the quality of data collected with Event History Calendar in comparison of data collected with Conventional Questionnaires (CQ) showed the superiority of the former in terms of reduction of underreporting and precision of locating the events in time (Belli, Shay, & Stafford, 2001; Belli, Smith, Andreski, & Agrawal, 2006). What it is yet not known is the performance of the EHC in regards to the seam effect problem. So far, mainly the dependent interviewing data collection technique has been shown successful in reducing the seam bias in a number of studies (Jäckle, in press).

The main hypothesis of this dissertation is that the EHC should reduce seam effects in comparison to CQ methods because it gives more retrieval cues to the respondent, it increases the precision of the location of the events in the timeline, and because it highlights contradictions, such as reporting being employed and unemployed at the same time. EHCs are better able to take advantage of combinations of memory retrieval strate-
gies simultaneously, including the use of landmarks and other contemporaneous events to anchor events on the timeline. Parallel probing (landmarks are a type of parallel probing) gives the respondents more retrieval cues because it takes advantage of the existence of interconnected thematic and temporal pathways that can be used to remember specific events (Belli, 1998). The flexible interviewing style of EHC allows respondents to retrieve events in the order with which they feel most comfortable. The structure of EHC highlights gaps and inconsistencies in the timeline (Caspi et al., 1996), especially in its computerized version (Belli, 2000), alerting the interviewer to probe for them thus possibly reducing item nonresponse and Don’t Know answers.

Two other papers discussed in section 1.4 gave further hints in the formulation of the main hypothesis. The first one is the use of a visual calendar containing information gathered during the previous wave as a memory aid. The strategy was shown effective in reducing seam effects (Kominski, 1990).

The second one is the paper by Rips, Conrad & Fricker (2003). The authors conceptualize seam effects as the contrast between memory for the most recent portion of one reference period, and memories for the most remote portion of the following reference period. Moreover “other things being equal, recall accuracy decreases with elapsed time” (Rips, Conrad, & Fricker, 2003, p. 527). Because recall error increases with time, respondents might use estimation strategies to compensate for retrieval difficulties. Estimation strategies tend to be less precise and they create higher inconsistencies at the seam than at the off-seam. Rips and colleagues advance the hypothesis that techniques such as EHC might be more successful in reducing the seam effect:

“Other techniques for reducing the seam effect might be more successful if they directly target forgetting, perhaps by giving respondents more retrieval cues than are available in a single conventional question. One such technique is the event history calendar in which respondents are encouraged to think about events from several contemporaneous life themes such as jobs, relationships, and household composition. With event history calendars, interviewers can press re-
respondents to align events across themes, and this can lead to additional recall” (p. 552).

4.1.2 Corollary hypotheses

Hill (1987) showed how an increase of the reference period between waves increased the magnitude of seam effect. Although this is the only paper reporting such an effect, it is plausible to think that seam effect increases with the increase of the reference period. Giving that recall accuracy decreases with time, the accuracy of the recall for the first month of the second reference period is likely to be lower for a two year wave length than a one year wave length thus generating higher seam bias.

The PSID interviews either the head or the wife asking about their job history. This design creates proxy answers in a household with both head and wife. If the head is interviewed the answers regarding the wife will be proxy and vice versa. This situation results in four possible combinations of answer at the joint of two waves as depicted by Figure 4.1.

<table>
<thead>
<tr>
<th>Head as target</th>
<th>Wife or ‘Wife” as target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave T</td>
<td>Wave T+1</td>
</tr>
<tr>
<td>Self</td>
<td>Self</td>
</tr>
<tr>
<td>Proxy</td>
<td>Self</td>
</tr>
<tr>
<td>Self</td>
<td>Proxy</td>
</tr>
<tr>
<td>Proxy</td>
<td>Proxy</td>
</tr>
</tbody>
</table>

Figure 4.1 Combinations of Head and Wife Answers at the Seam Creating Four Self/Proxy Reports

Proxy answers can potentially have an effect on the magnitude of the seam bias. As Moore and Kasprzyk (1984) have shown, the magnitude of the effect at the seam for the four combinations of self and proxy answers (Self-Self, Self-Proxy, Proxy-Self, and
Proxy-Proxy) depends on the kind of variables used in the analysis. Alwin (2007) was able to measure the reliability of answers of proxy reports (in this case the spouse) for items that were placed in the same questionnaire at a different place. Proxy reports were slightly less reliable than self reports for variables such as years of schooling, occupational status and hours worked per week. Given these premises it will be interesting to study the impact of proxy reports on the amount of seam bias taking into account the four combinations that are created at the seam. Self and proxy results will be studied in a multivariate model controlling for other variables that might have an effect on the magnitude of the seam such as age, education, sex, race, time elapsed from the interview, and number of job transitions within a year. Because validation data are not available in the data-set being analyzed, the study of the impact of proxy reports on the seam bias will be exploratory in nature and some assumptions about the distribution of the data will have to be made in order to proceed, as detailed in section 4.4. The problem is to distinguish real transitions from spurious transitions at the seam. In other words, the results from the analysis will give some point estimates at the seam but in absence of validation data of some sort (such as the test-retest design implemented using overlapping waves in the HILDA) it is not possible to “attach” a variable to each subject telling if the transition at the seam is “real” or a reporting error.

4.1.3 Expected findings

From the above description, and the findings from the literature summarized in chapter 1, the following outcomes are expected to be found in the data:

- Labor status changes will display seam bias for the transition between December and January of two waves.
- Seam bias for labor force status changes collected with EHC two year reference period should be of lower magnitude than the CQ two year reference period.
The move to a two-year data collection begun in 1997 should increase the magnitude of CQ seam effects in comparison to the previous CQ one-year waves. At this stage there are no sufficient data and knowledge to expect if the magnitude of seam bias for CQ one year reference period would be lower or higher than the EHC seam bias.

Proxy reports creates four combinations of answer at the seam that might have an impact on the magnitude of bias. At this stage it is not possible to hypothesize the direction of any effect.

4.1.4 Data analysis plan

Following the expected findings, the first step will be to look at the magnitude of the labor status changes for each seam point (see figure 4.2) and compare them with the status changes between November to December of the previous year (time $T-1$). The November-December status changes point is the closest to the interview date and for this reason considered the most reliable. Statistical tests taking into account the dependency of the data will be calculated for each pair of seam vs. November-December comparisons.

The second step will be to compare the magnitude of the 2 year CQ seam point with the 2 year EHC seam point.

The third step will be to compare the magnitude of the remaining seam points among each other. Of most interest will be the comparison of the one year recall period seam effect with the two year recall period seam effects.

Lastly, two logistic regression models, one for the 2 year CQ and the second one for the 2 year EHC seam point will be estimated to predict group membership at the seam (1 change at the seam, 0 no change at the seam). Predictors variables based on previous studies and hypotheses generated during the data analysis will be included.
4.2 Data and methods

To test the hypotheses of this study, a concatenated dataset for the 1995-2005 waves of the Panel Study of Income Dynamics (McGonagle & Schoeni, 2006; PSID staff, 2006a) was constructed. More specifically, the dataset for waves 1995-2001 was obtained from the PSID data center, an online database that enables the user to create a customized subset and companion documentation of the public release data (Chiteji & Stafford, 2004). PSID collected data annually until 1997, at which time the panel moved to a biennial data collection. For biennial waves the PSID data center makes available data referring to events that happened one year before the interview, also called time $T-1$ (see figure 4.2 later on). Data referring to time $T-2$ (example: for the 1999 wave data referring to calendar year 1997) are stored in separate files called 1999 and 2001 $T-2$ individual income files and they are not part of the online database but can be found in the supplemental PSID data section of the website. The 1999 and 2001 $T-2$ dataset were then merged with the 1995-2001 dataset. For the last two waves, 2003 and 2005 ($T-1$ and $T-2$ data), a pre-release customized dataset was made available by the PSID principal investigators and merged into the previous file. Although in the EHC the data are collected with a third of a month level of detail (chapter 3), only month variables were made available for this analysis to model the level of analysis available in the CQ waves. At the time of this writing only the EHC 2003 $T-1$ official release wave is available to the public data user.

---

8 http://simba.isr.umich.edu/
10 University of Nebraska Lincoln IRB approval code 296.
11 The 2003 T-1 public release data is however available on the Data Center website. For consistency reasons it was chosen to use the 2003 -2005 T-1 and T-2 pre-release data. A test of differences between the 2003 T-1 public release and the pre-release data for the same period revealed that the former contained slightly more cases than the latter (5%).
The variables used for the analyses are labor force questions aimed to measure monthly labor force status for the head and wife of the household. Figure 4.2 shows the waves used in the analyses and the reference period of each wave.

Figure 4.2 PSID Waves Used in the Analysis and Seam Points.
Note: WWS = within-wave seam.

Figure 4.2 also indicates the methodology of data collection, CQ or EHC, and seam points. Seams 1 to 6 are the standard seams that occur when joining two waves of data collected at different years. The within-wave seam points (WWS) became available when the PSID started collecting data referring to a two year reference period and are referred to the transitions between December of the first year (T-2) to January of the second year (T-1). In total, there are ten seam points that will be the objects of the analyses. More information about the sampling design, response rates, and the survey content are found in McGonagle and Schoeni (2006).

For the waves that are the object of this analysis (1995-2005), data were collected using computer-assisted telephone interviewing, partially from a centralized location and partially from the interviewer’s homes. In-person interviews have been collected only with respondents who do not have a phone or where telephone interviewing is problematic. In the 1999 wave, for example, 97.5% of the data were collected over the phone. Beginning in 2003, a computer–assisted Event History Calendar instrument (Belli, 2003)
was integrated with the current CATI instrument (Blaise) and major sections of the questionnaire were administered that way.

4.2.1 Question wording for labor force status

PSID collects labor force status data only about the Head and the Wife of each household. The questionnaire contains separate questions for the Head and the Wife. Because only one respondent is selected for the interview, the answers for the Head section could be self or proxy depending on who is answering and vice versa. PSID attempts to interview either the Head of the Wife, other household members are rarely used to provide proxy reports.

The CQ questionnaire was the same for the waves 1995 to 1997 and 1999 - 2001 for events that happened during the year before the interview, time $T-1$. Figure 4.3 shows a flow chart with the key variables and their question wording used in the dataset.
Figure 4.3 CQ Questionnaire Flowchart for Labor Force Variables 1995-97 and 1999, 2001 T-1.

Since the questionnaire is the same for the Head or the Wife, in order to provide a more straightforward account, I will present explanations using the Head as an example. The employment section starts with a main branching pattern: if the Head is working during the time of the interview, section B of the questionnaire is asked, otherwise section C is asked. For each of the three labor force states, employment, unemployment, and not in the labor force, a question is asked referring to the previous year in order to ascertain in which month(s) the respondent was working, looking for a job, or unemployed and not looking for a job. This procedure produces 12 variables (one for each month) in the dataset for employment, 12 for unemployment, and 12 for not in the labor force, for each sec-
tion of the questionnaire, B and C for the Head, and D and E for the Wife. After the question about the employment status (B/D39), detailed questions about each job (main and secondary) were asked such as a description of the occupation, duties, kind and name of business. Before asking about unemployment and out of labor force status, questions about work missed because of sickness, vacation and strike were asked (B/D60-B/D/71). In the appendix, a full flow chart with most of the labor questions referring to Head is reported. Complete questionnaires are available at the PSID main data, documentation and questionnaires webpage.\(^\text{12}\)

When the PSID switched to a two-year data collection in 1997, the questions about the job status referring to two years before the interview \((T-2)\) were asked in a more simplified way, and not consecutively after the question referring to time \(T-1\). Moreover, the “not in the labor force” (N) question was not asked for \(T-2\).

\(^{12}\) [http://simba.isr.umich.edu/Zips/ZipMain.aspx](http://simba.isr.umich.edu/Zips/ZipMain.aspx)
In 2003 the PSID switched sections of the questionnaire to computerized Event History Calendar. The labor force sections were part of the switch. Scripts for the EHC are reported in Figure 4.5.

Figure 4.5 Questionnaire Flowchart for EHC Labor Force Variables 2003 and 2005 $T-1$, $T-2$

There are many differences in comparison to the CQ data collection. Besides the entire EHC interviewing style, the scripts of the questions are different. First of all, the
questions are referred to the previous two years and information about time $T−2$ is asked in the same section and not later in the questionnaire as for the 1999 and 2001 CQ waves. Second, the wording of the scripts in the EHC is different. Instead of asking in which month the respondent was employed, unemployed or not in the labor force, it is asked when either of those events happened (*When was that?*). Third, questions regarding “not in the labor force status” were asked before the unemployment questions and not after as in the CQ. Last, questions regarding a detailed description of the job, and work missed because of sickness, vacation and strike were asked after the E, U, N status questions and not before as in the CQ. It is however true that due to the flexible interviewing style of EHC we cannot be sure about the order of the questions since the interviewers are allowed to follow the respondent’s way of recollecting events. The EHC computerized calendar was also programmed to check on mutual exclusiveness of labor force status. The differences in question order pointed out here are based on a strict comparison between CQ questionnaire and EHC script. In the results section these differences will be further considered.

4.2.2 *Variables*

For the CQ questionnaire, the first step was to combine answers from the sections (B, C) for the head together and (D, E) for the Wife together. In fact, the first filter question in the CQ questionnaire (head or wife is working now, see Figure 4.3) creates two sets of variables for the same reference period. Even if the respondent answered that he/she is not working now, he/she might have worked during the reference period anyway (interviews start in February and end late November).

In the EHC questionnaire there was no need to combine values of employment (E), unemployment (U) and not in the labor force (N) stored in different variables as shown in Figure 4.5. It is important however to notice how the values for E, O, and N were converted from a third of a month detail to month values. One of the features of the PSID EHC instrument is to record events with a third of a month level of detail (see
chapter 2). Third of a month information is then converted into month information in the public release dataset. Specifically for labor force status the conversion was done combining the three thirds of a month E, U, N in a single month. Table 4.1 shows an example.

Table 4.1
Examples of EHC Conversions from Thirds of a Month to Month String for Labor Force Status

<table>
<thead>
<tr>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Case 4</th>
<th>Case 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>I I I</td>
<td>I I I</td>
<td>I I I</td>
<td>I I I</td>
<td>I I I</td>
</tr>
<tr>
<td>E E E</td>
<td>E U U</td>
<td>N E E</td>
<td>U U U</td>
<td>U N U</td>
</tr>
</tbody>
</table>

New variables created:
- Employed month
- Unemployed month
- Not in the labor force month

In order to streamline the data analysis, the entire dataset was restructured from a household level to a person level. If a respondent was Head for that wave, the Head job status records were associated with him/her and vice versa for the case in which a person was classified as Wife for that wave. The major advantage of this method is that the person level file solved of the problem of the switch between Head and Wife (and vice versa) among waves\(^{13}\), crucial for the computation of the seam transition.

Three variables were constructed for each month containing information about employment (e.g. E= person was employed that month), unemployment and not in the labor force. Employer changes were not taken into account. A person was classified as working in that month taking any of the six possible jobs (main job, secondary job and so on) that were allowed to be recorded in the CQ instrument or any of the eight possible jobs that were allowed to be recorded in the EHC instrument. The month to month status

\(^{13}\) A classic example is when a female Head, for example living single at one wave, gets married and, due to the PSID definitions, becomes automatically Wife at the next wave.
changes variables were created concatenating the three variables for month $m$ with the three variables for month $m+1$. Table 4.2 exemplifies some CQ possible cases.

Table 4.2
Examples of Possible CQ Status Changes

<table>
<thead>
<tr>
<th></th>
<th>Month $m$</th>
<th>Month $m+1$</th>
<th>Transition point $m_m+1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person 1</td>
<td>E</td>
<td>U</td>
<td>EU</td>
</tr>
<tr>
<td>Person 2</td>
<td>U</td>
<td>U</td>
<td>UU</td>
</tr>
<tr>
<td>Person 3</td>
<td>N</td>
<td>UE</td>
<td>NUE or NEU</td>
</tr>
<tr>
<td>Person 4</td>
<td>EU</td>
<td>E</td>
<td>EUE or UEE</td>
</tr>
</tbody>
</table>

Person 1 reports having lost his/her job at month $m+1$, becoming a mover for that transition. Person 2 is a stayer, he/she reports being unemployed at month $m$ and at month $m+1$. Person 3 moved from being out of labor force to being unemployed and the employed in the same month $m+1$. This person probably was not looking for work at month $m$, started looking at the beginning of month $m+1$ and then found a job in the same month. Person 4 reports being employed and unemployed in the same month, possibly because the two events happened in different weeks.

The table also shows the problematic nature of having multiple transitions per month. The problem with multiple transitions in the CQ questionnaire is that it is not possible to assess the temporal order because the data are on a month level of detail. If the respondent reports to be employed and unemployed in the same month, there is no way to know if this person was employed and then unemployed or unemployed and then employed. In the case of person 3 of table 4.2, for example, the transition can be either NU or NE depending on which order E and U happened at month $m$. As discussed earlier, multiple transitions in the same month were not considered for the analysis. Multiple transitions within the month are theoretically avoidable in the EHC because the data are collected originally on a third of a month level of detail. This level of detail is unfortunately lost when the data are released to the general public in order to make the variables comparable with the previous waves. Transitions such as of person 3 and 4 are thereby called *unclassifiable* and excluded from the analysis.
4.2.3  *Self and proxy answers: a possible confounding factor*

Given that the PSID uses proxy interviews to collect labor force history from the Head or the Wife, it might be necessary to control for it during the data analysis. Even if the proxy is almost always the partner, it is very possible that the proxy information about the job history is not as accurate as a self answer. It is important to remember however that the PSID “one year rule” (PSID staff, 2005b) applied for the non legally married couples (living together for at least one year during time \( T - 1 \) in order to be qualified as “Wife” otherwise the person is classified as girlfriend and boyfriend with the consequence to not collect labor force history for him/her ) makes the “Wife” no ordinary proxy but possibly very knowledgeable about the partner’s work history. When the PSID switched to a biennial data collection the “one year rule” remained the same possibly making the “wife” of biennial waves less knowledgeable about time \( T - 2 \).

The self/proxy issue creates also another problem when analyzed in the PSID data. For example it is possible that during one wave answers about the self are reported and then at the next wave answers about this individual are reported by the spouse as proxy reports. When computing transitions rates at the seam, the self/proxy issue creates 4 possible combinations (self-self, self-proxy, proxy-proxy, proxy self) as previously shown in Figure 3.3. In Figure 4.6 the mixture of the four combinations of self or proxy answers is reported for each seam point.
Figure 4.6 Mixture of the Four Combinations of Self or Proxy at each Seam, Percentage of Answers

From the figure it appears that there are not large variations in the mixture of self and proxy answers combinations at the seam. If this is the case, self and proxy answers can be combined in an unique analysis with no need to treat them separately. In order to test if the difference in mixture of self and proxy answers for pairs of seam points reaches statistical significance I conducted the Marginal Homogeneity test. The test of Marginal Homogeneity (Agresti, 2002) accounts for the fact that the answers of panel respondents are dependent and for the multinomial outcome of ordinal data (4 possible combinations). The null hypothesis states that the row and column marginal response distribution of the four self-proxy combinations will be the same for any two seam point (Sheskin, 2004).
The alternative hypothesis states that for at least one combination, the marginal distribution for one seam point will not be equal to the marginal distribution for another seam point. The test is performed only with the subjects whose answers are present in both transitions and the results are presented in the following table.

Table 4.3
Test of Marginal Homogeneity to Assess Differences in the Mixture of Self or Proxy Answers Among Different Seam Points

<table>
<thead>
<tr>
<th>Seam comparison</th>
<th>St. MH Stat.</th>
<th>Asymp. Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQ Seam 95_96 vs. CQ Seam 96_97</td>
<td>-1.057</td>
<td>.291</td>
</tr>
<tr>
<td>CQ Seam 98_99 vs. CQ-EHC Seam 00_01</td>
<td>1.078</td>
<td>.281</td>
</tr>
<tr>
<td>CQ-EHC Seam 00_01 vs. EHC Seam 02_03</td>
<td>.388</td>
<td>.698</td>
</tr>
<tr>
<td>CQ Seam 98_99 vs. EHC Seam 02_03</td>
<td>1.632</td>
<td>.102</td>
</tr>
</tbody>
</table>

Four pairs of seam points are tested and none of them differ in a systematic way. The reason for the above four comparison is to compare seam points that are the closest as possible in time, to make the comparison meaningful.

Self-proxy answers at the seam also present a potential confounding factor. In a household where Head and Wife are present, it is possible that there is a correlation of error between the self and the proxy answers. If this is the case, self answers from a single Head household could not be treated in the same way as self answers from a Head – Wife household. For this reason, a correlation between self and proxy answer (classified as staying (EE, UU, NN) or moving (EU, EN, UE, UN, NE, NU) was computed for households in which answers for the Head and the Wife were present. Because at the seam there might a switch between self and proxy answers as shown in Figure 3.3, two Pearson’s $r$ correlations were computed\(^\text{14}\), the first one is where there is no switch of respondents at the seam (self self or proxy proxy), the second one when there is a switch

\(^{14}\)Strictly speaking, the Phi ($\phi$) coefficient should be used when computing correlation between two dichotomous variables. However the phi coefficient and Pearson’s $r$ are algebraically equivalent when computed on dichotomous variables.
(self proxy and proxy self). Results from the correlations are reported in Table 4.4. Values of $r$ for both EHC and CQ indicate a very small but significant correlation between self and proxy answers when the respondent is the same across waves. The correlation between self and proxy answers when there is a switch of respondent between waves is not statistically significant. These results provide evidence in favor of treating answers for single household together with answers from a Head-Wife households.

Table 4.4
Pearson $r$ Correlations Between Self and Proxy Answers from the Same Household

<table>
<thead>
<tr>
<th>Sample</th>
<th>$r$</th>
<th>Approx. sign.</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQ 1998_99 Same respondent in both waves</td>
<td>.106</td>
<td>.000</td>
<td>2,386</td>
</tr>
<tr>
<td>CQ 1998_99 Switch of respondent between waves</td>
<td>-.010</td>
<td>.908</td>
<td>135</td>
</tr>
<tr>
<td>EHC 2002_03 Same respondent in both waves</td>
<td>.042</td>
<td>.031</td>
<td>2,689</td>
</tr>
<tr>
<td>EHC 2002_03 Switch of respondent between waves</td>
<td>-.022</td>
<td>.783</td>
<td>159</td>
</tr>
</tbody>
</table>

Table 4.4 suggests an even smaller correlation for EHC answers in comparison to CQ answers among the same respondents across waves. A possible research question would be to test if the EHC correlation is significantly lower than the CQ correlation. Due to the dependency of the correlations because most of the same subjects are respondents in the two waves, it is necessary to use a test that accounts for dependent correlations (Chen & Popovich, 2002). Steiger (1980) proposes using Dunn and Clark (1969) $z$ test formula to test for the difference between two dependent correlations with zero elements in common for samples > 20. Dunn and Clark $z$ test cannot be readily computed in the majority of statistical packages. The results were then obtained using Depcor, a Fortran 77 program for comparing dependent correlations (Silver & Hittner, 2006). Due to the small sample size of the CQ and EHC correlation when the respondents are switching between waves (line two and four of Table 4.4), the Dunn and Clark $z$ test was not computed. When comparing the two correlations between CQ (.106) and EHC (.042) the ob-
tained Dunn and Clark\textsuperscript{15} $z$ is of 2.097 with $p$ value of 0.018. This means that the correlation of error between self and proxy responses in the CQ is higher than for the EHC.

4.2.4 *Data treatment, imputation and issues with some variables*

In order to make the waves as comparable as possible, a subsample of Latino households was excluded for the 1995 wave. The reason is because the Latino sample, around 2000 families, was added in 1990 but dropped after 1995. In 1997, a new sample of immigrants was introduced in the study, starting with 441 families in 1997 and reaching 511 in 1999. This sample was dropped from the analysis to keep the panel as homogeneous as possible for the seam effect analysis which includes the 1995 and 1996 waves. In 1997 the PSID also reduced the core sample from the nearly 8,500 families in 1996 to approximately 6,168 in 1997 as shown in Figure 3.1 (McGonagle & Schoeni, 2006).

In the 1999 and 2001 waves, a question about “not in the labor force status” (N) was not asked for time $T-2$. In order to solve this problem, an imputation strategy has been applied to compute the “N” variable necessary to calculate the transitions. A “N” status was imputed for the months in which the respondent reported to be retired. If the panel participant reported to be working (not retired) during the year time $T-2$, the N status was imputed in the month where the respondent was both not working and not looking for a job in the same month. Results from the CQ 1999 and 2001 $T-2$ transitions should be then interpreted with caution.

The variable regarding unemployment for February 1994 of the Wife is missing from the official release dataset. The computations regarding the reference period 1994 start then from March.

In January 1997, an abnormal number of people, approximately 3000, reported unemployment for that month against an average month unemployment status of 340 during the same year. The causes for this abnormality are still not known, but this number

\textsuperscript{15} Since the formula asks for only one sample size and the N was different for any pair of correlation due to attrition and item nonresponse, the harmonic mean of 2,127 was used.
could not have been used for the analysis, especially because it was impacting the seam calculation (seam 3, Figure 4.2). In order to make the computation for the transition possible, the number of people reporting being unemployed in February of 1997 was used as surrogate variable for January of 1997. This abnormality was found when data started to be collected for a reference period of two years for the first time. The change from one to two years reference period could have resulted in a CATI programming glitch that created the abnormal number of people resulting unemployed for January of 1997. The problem in fact did not present itself in the 2001 wave.

A similar phenomenon happened in the first wave of EHC data collection. An increased number of people, compared to the previous waves, reported having two or more transitions within the same month. The most notable was reporting being EN (or NE). The approximate number of people being EN was 340 against an average of 25. This combination was excluded from the computation because unclassifiable but it is likely that some of these case would be classified otherwise. The causes for these irregularities are still not known but they coincided when the PSID collected data with EHC for the first time. Interestingly, the problem disappeared in the second wave of EHC data collection (2005) as shown in Table 4.5 where the number of unclassifiable transitions went back to values of the previous CQ years.

Lastly, before reporting the main results and in order to make meaningful comparison across seam points, the average age at the seam has to be investigated. If for example the panel is aging the comparison across seam points are at stake because age can have an impact on the mixture of job transitions. The PSID following and eligibility rules however avoids the aging of the panel across waves. In fact, the mean age for the six seam points is not really moving in any meaningful direction for the 10 years object of investigation (around 43.5 years of age).
4.3 Primary findings

In order to answer the main question of this dissertation (is the magnitude of seam effect reducing when PSID switched to EHC?) the nine classifiable labor force status changes (Table 1.1) were calculated from March-April 1994 to November-December 2004. When reporting results, the names of the variables will be referred by the reference period investigated in each wave, and not to the year in which the data were collected. This will simplify the interpretation and clarify the seam points (see Figure 4.2). Table 4.5 provides the average within wave number of the nine possible classifiable status changes and the percentage of unclassifiable status changes (see table 4.2 for an explanation) among all possible status changes (base).

Table 4.5.
Number of the Nine Classifiable Status Changes and Percentage of Unclassifiable Status Changes for Each Reference Period and at each Seam Point

<table>
<thead>
<tr>
<th>Reference Period</th>
<th>N of nine classifiable status changes</th>
<th>% of unclassifiable status changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>12,278</td>
<td>1.86</td>
</tr>
<tr>
<td>CQ Seam 1994_95</td>
<td>11,408</td>
<td>2.22</td>
</tr>
<tr>
<td>1995</td>
<td>12,399</td>
<td>1.79</td>
</tr>
<tr>
<td>CQ Seam 1995_96</td>
<td>8,535</td>
<td>2.65</td>
</tr>
<tr>
<td>1996</td>
<td>9,234</td>
<td>1.72</td>
</tr>
<tr>
<td>CQ Seam 1996_97</td>
<td>7,427</td>
<td>2.35</td>
</tr>
<tr>
<td>1998</td>
<td>9,511</td>
<td>1.41</td>
</tr>
<tr>
<td>CQ Seam 1998_99</td>
<td>8,740</td>
<td>1.72</td>
</tr>
<tr>
<td>2000</td>
<td>10,312</td>
<td>1.38</td>
</tr>
<tr>
<td>CQ EHC Seam 2000_01</td>
<td>8,648</td>
<td>5.29</td>
</tr>
<tr>
<td>2002</td>
<td>10,083</td>
<td>2.50</td>
</tr>
<tr>
<td>EHC Seam 2002_03</td>
<td>8,919</td>
<td>4.97</td>
</tr>
<tr>
<td>2004</td>
<td>10,432</td>
<td>1.43</td>
</tr>
</tbody>
</table>

Note. The N is an average for the year (January-November), and exact at the seam. The data have been adjusted for the abnormal number of EN transitions during the first EHC data collection as discussed in the previous section.
The number of nine classifiable status changes represents roughly the number of answers regarding Head and Wife labor force status. It is slightly lower than the total numbers of Head and Wife answers in the panel due to item nonresponse and unclassifiable status changes within a month. The percentage of unclassifiable status changes is very small and relatively stable across the waves making comparisons across seam points more consistent. However, an increase in the percentage of unclassifiable status changes (more than one status change in the same month) at the seam is noticeable. When the PSID switched to EHC the percentage of unclassifiable status changes at the seam increased as well (5.29% for seam 2001_01 and 4.97% for seam 2002_03).

Figure 4.7 shows transitions of the movers for the period under investigation. The chart is created plotting for each pair of months the percentage of movers for the six categories in which status changes could take place (EU, EN, UE, UN, NE, NU) among all the stayers and movers nine classifiable status changes (see table 1.1). For example the EN percentage for the transition point January to February is obtained dividing the number of reports being E in January and N in February (EN) divided by the reports (EE + EU + EN + UE, + UU + UN + NE + NU + NN)*100. Multiple status changes per month (unclassifiable) are not reported for the reason described above. The chart is built with unweighted data. A chart plotted with weighted data using the PSID person level weights up to the 2003 wave showed a very similar pattern (results not shown). It is not possible to plot a chart with weighted data because the 2005 wave weights are not ready at the time of this writing.
Figure 4.7 Month to Month Status Changes of Movers in Waves 1995 – 2005, All Answers, no immigrant Samples.
For better interpretation, the above picture should be printed in color.
The description of the findings with the appropriate statistical testing will start with a comparison of each seam point with the closest within-wave transition (November December) to test if transitions at the seam are real bias and not a fluctuation in the data. The second comparison will be conducted between the two year CQ seam (1998_99) and the two year EHC seam (2002_03) in order to support or reject the main hypothesis of this dissertation i.e. that the magnitude of the EHC seam bias is lower than the CQ seam bias. Then other seam points will be compared among themselves (example CQ one year with CQ two years). Then within-wave seam effect will be discussed. Lastly, the focus will be on the transitions with the highest seam effect.

4.3.1 Comparison of seam points with previous November-December status changes

As discussed at the beginning of this chapter, seam effects for labor force status changes are not documented in the PSID. To test if seam status changes are of a magnitude to be considered a problem or if they are just a minor source of noise in the data (Willis, 2001), seam points will be compared with the immediate previous status changes of November to December. The basis of this reasoning is conceiving the November-December status changes as the most reliable because closest to the time of the interview. The second reason is because if there are seasonal fluctuations in the job market, it makes sense comparing the December-January status changes to the closest point in time: November to December. Results from the tests in Table 4.6 confirm the initial hypothesis that seam bias is not a minor source of noise in the data but a phenomenon considerable magnitude.
Table 4.6
Test of Marginal Homogeneity Between Seam Points and Immediate Previous Status Changes November-December

<table>
<thead>
<tr>
<th>Comparison</th>
<th>St. MH Stat.</th>
<th>Asymp. Sig. (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seam 1995_96 vs. Nov-Dec 1995</td>
<td>3.933</td>
<td>.000</td>
<td>8,455</td>
</tr>
<tr>
<td>Seam 1996_97 vs. Nov-Dec 1996</td>
<td>2.515</td>
<td>.012</td>
<td>7,639</td>
</tr>
<tr>
<td>Seam 1998_99 vs. Nov-Dec 1998</td>
<td>8.779</td>
<td>.000</td>
<td>8,669</td>
</tr>
<tr>
<td>Seam 2000_01 vs. Nov-Dec 2000</td>
<td>7.981</td>
<td>.000</td>
<td>8,579</td>
</tr>
<tr>
<td>Seam 2002_03 vs. Nov-Dec 2002</td>
<td>5.803</td>
<td>.000</td>
<td>8,808</td>
</tr>
</tbody>
</table>

4.3.2 Comparison of CQ and EHC magnitude of seam points

From Figure 4.7 it appears that the magnitude of the two year CQ seam (1998_99) appears higher than the EHC two year seam bias (2002_03). In order to test if this difference reaches statistical significance the test of marginal Homogeneity is applied on the subjects who were present at both waves during the panel (listwise deletion). Results are presented in table 4.7.

Table 4.7
Test of Marginal Homogeneity Between CQ Seam and EHC Seam

<table>
<thead>
<tr>
<th>Comparison</th>
<th>St. MH Stat.</th>
<th>Asymp. Sig. (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQ 1998_99 vs. EHC 2002_03</td>
<td>8.449</td>
<td>.000</td>
<td>6,559</td>
</tr>
</tbody>
</table>

Results from the test support the initial hypothesis of this dissertation: EHC data collection methodology reduced the seam effect in comparison to the CQ. In comparing CQ and EHC seam points we have to remember the following. Because the test of Marginal Homogeneity uses the same subjects across waves, the status changes of the EHC seam points belong to subjects slightly older than when the measurement was computed in the CQ. This might be a potential confounding factor because older subjects can have different status changes patterns. Second, although in the analysis it was possible to con-
trol for some confounding variables such as the immigrant sample, others remained present: the comparison of EHC and CQ seams is confounded by the wording of the questionnaire. In the CQ case, people were asked about employment, unemployment and not in the labor force with months as response options (i.e. in which months during [previous year] were you working for [name of employer]). In the EHC case, the question wording required more precision in remembering the job history (e.g. when did you start and stop working for [name of employer]). When thinking about these events with month as a unit of measurement in the CQ, the respondent could simplify the remembering thus possibly reducing seam effects. In addition, data might be affected by order effects; in the previous discussion of the CQ and EHC questionnaires (Figures 4.3 to 4.5) it was noticed how the N question was asked before the U question in the EHC and the other way around in the CQ. Also, in the CQ, specific questions about the job and time missed for sickness, vacation and strike were asked after the employment section whereas in the EHC they were asked after the entire E, N, U section. Asking these questions after the timing of E, U, and N could have given less retrieval cues to the respondents although in the EHC the interview is more flexible and it is easier to go back and forth on the timelines making adjustments accordingly to the way the respondent retrieves information. All these differences in question wording make the comparison between CQ and EHC problematic at best. The present study does not include a control group in which question wording was asked in CQ mode or EHC mode as similarly as possible.

4.3.3 Comparison of one year recall period seam magnitude and two year recall period seam magnitude

Based on the above hypothesis, it is expected that the one year recall period CQ seam point would be of lower magnitude than the two year recall period seam point. In other words we are expecting a statistically significant difference between the one year CQ and the two years seam points. In order to test the above hypothesis the first seam point 1994_95 will be compared with the two year seam points starting from the 1998_99 CQ.
Because the two year seam points are far away in time from the 1994_95 seam, it is not fair to consider only the same cases present at both waves because the same subjects are getting older this having a different job history pattern than when they were younger. This strategy will also reduce the available sample size. The test will then consider the two samples of each seam point as independent. Table 4.8 present results of three Chi square comparisons among the 1994_95 one year recall period and the three seam points with a two year recall period.

Table 4.8
Chi Square Test Between One Year Seam Point and Two Year Seam Points

<table>
<thead>
<tr>
<th>Comparison</th>
<th>$X^2$</th>
<th>$p$</th>
<th>Df</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 year CQ 1994_95 vs. 2 year CQ 1998_99</td>
<td>117.01</td>
<td>.000</td>
<td>8</td>
</tr>
<tr>
<td>1 year CQ 1994_95 vs. 2 year CQ_EHC 2000_01</td>
<td>124.81</td>
<td>.000</td>
<td>8</td>
</tr>
<tr>
<td>1 year CQ 1994_95 vs. 2 year EHC 2002_03</td>
<td>155.84</td>
<td>.000</td>
<td>8</td>
</tr>
</tbody>
</table>

Each comparison reaches statistical significance, meaning that the one year seam bias is always lower than any two year seam bias. These results provide further evidence to the initial finding by Hill (1987) that increasing the recall period increases the amount of seam bias. Hill used two panels (PSID and SIPP) to come to this conclusion while with this dataset it was possible to test the above hypothesis using the same panel thus making the conclusion stronger. Because the one year CQ seam bias is significantly lower than the two year EHC seam bias we can also infer that the EHC methodology was not able to offset the increase of seam bias due to the increase of the recall period.

4.3.4 Comparison of two year recall period seam points

In the next table the two year seam points are compared among each other using the test of marginal homogeneity.
Table 4.9
Test of Marginal Homogeneity Between Two Year Recall Period Seam Points

<table>
<thead>
<tr>
<th>Comparison</th>
<th>St. MH Stat.</th>
<th>Asymp. Sig. (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 year CQ 1998_99 vs. 2 year CQ-EHC 2000_01</td>
<td>3.934</td>
<td>.000</td>
<td>7,118</td>
</tr>
<tr>
<td>2 year CQ-EHC 2000_01 vs. 2 year EHC 2002_03</td>
<td>6.665</td>
<td>.000</td>
<td>7,527</td>
</tr>
</tbody>
</table>

The comparisons reach statistically significance indicating a reduction of seam bias from the CQ to the hybrid seam (CQ-EHC) to the EHC (lowest among the two year reference period seam points).

4.3.5 *An unexpected phenomena: within-wave seam bias*

When looking at Figure 4.7, previously undocumented phenomena appear in the data: the presence of within-wave seam effects when the PSID switched to biennial data collection, for example the CQ_WWS_97_98 depicted also in figure 4.2. Movers status change between December of the first year and January of the second year of the reference period appears higher than in November and December of the first year, or January and February of the second year. The effects seem surprising at first, because the data were collected during the same interview. On the other hand, in the CQ instrument, respondents were encouraged to report labor force status for time $T^*_{−2}$ separately from time $T^*_{−1}$. At the beginning of the questionnaire $T^*_{−1}$ job status questions were asked and later in the questionnaire, more or less 30 minutes later, $T^*_{−2}$ questions were posed. The $T^*_{−2}$ question wording regarding job status (figure 4.4) was also much more simplified as discussed in section 4.2.1. On the other hand the within-wave EHC seams (e.g. EHC_WWS_01_02) are however nonexistent. In fact questions were asked concurrently referring to the two year reference period for E and N status, and for $T^*_{−1}$ and then $T^*_{−2}$ for the U status (see Figure 4.5). This means that when the PSID moved to EHC the change in the design of the questionnaire solved the within wave seam effect found in the CQ. Results from statistical test presented in Table 4.10 support the conclusion drawn from the visual inspection of Figure 4.7.
Table 4.10
Test of Marginal Homogeneity Between Within Wave Seam Points and Immediate Previous Transitions November-December

<table>
<thead>
<tr>
<th>Comparison</th>
<th>St. MH Stat.</th>
<th>Asymp. Sig. (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQ WWS Seam 1997_98 vs. Nov-Dec 1997</td>
<td>-5.343</td>
<td>.000</td>
<td>8,246</td>
</tr>
<tr>
<td>CQ WWS Seam 1999_00 vs. Nov-Dec 2000</td>
<td>-5.966</td>
<td>.000</td>
<td>9,570</td>
</tr>
<tr>
<td>EHC WWS Seam 2001_02 vs. Nov-Dec 2001</td>
<td>-.567</td>
<td>.571</td>
<td>10,062</td>
</tr>
<tr>
<td>EHC WWS Seam 2003_04 vs. Nov-Dec 2003</td>
<td>.203</td>
<td>.839</td>
<td>10,368</td>
</tr>
</tbody>
</table>
4.3.6  *Labor force status changes with the highest seam bias*

The reader would have been noticing how the EN (purple line) and NE (green line) transitions are more sensitive to seam effect. In order to better visualize them, Figure 4.8 plots only EN and NE transitions.

Figure 4.8. Month to Month Transitions of EN and NE Movers in Waves 1995 – 2005, All Answers, no Immigrant Samples.

An explanation for this phenomenon might be in the difficulty for the respondent to separate the concepts of “unemployment” from “not in the labor force” that, although clear in the official definition, have been proven to be of difficult comprehension for the respondent (Campanelli, Martin, & Rothgeb, 1991). Another cause could be the fact that it is more difficult to remember a *non event* (was there any time when you were unemployed and *not* looking for a job) than an event. These results are similar to the first
analysis of labor force transition in the SIPP by Martini (1989). He found the highest ra-
tio seam vs. off-seam for the transitions EN, NE and also for the transition UE (see Fig-
ure 1.8).

4.4 Secondary findings

The second main question originated from the dissertation proposal was to explore which
person level variables (e.g. age, sex, education, race) or design level variables (e.g. self or
proxy answer, including time lag between the interview and the reference period, and
having the same interviewer at both waves) will have an effect on the magnitude of seam
effect for the CQ two year reference period and the EHC two year reference period. In
the literature previously reviewed in Chapter 1, these studies contained some kind of
validation data (e.g. reinterview or external record) that were used to disentangle real
changes at the seam from spurious one-seam effect (D. H. Hill, 1987; Jäckle & Lynn,
2004). In fact some people have a legitimate transition at the seam that it is hypothesized
to be comparable to previous within-wave month to month transitions. The small number
of studies reviewed reveals how uncommon is to have validation data that will enable to
precisely pinpoint which subject had a real change and which did not at the seam. The
data used in this dissertation are no exception: for the seam point CQ 1998_99 and the
EHC 2002_2003 there are no validation data to be used as reference for comparison pur-
poses. The only PSID exceptions are the mentioned study by Hill (1987) with reinterview
data and the study by Belli, Shay and Stafford (2001) where a sub sample of the PSID
participants (N=616) was reinterviewed in 1998 regarding events happened in 1997 and
1996.

Because validation data are not available, the strategy used here is to treat all
movers status changes at the seam “as seam bias”. The logic is that even if we know that
some people move legitimately at the seam, we also know that the majority of transitions
at the seam are error. A simple calculation can give us an idea of the magnitude of this
error. Assuming that the sum of November to December movers status changes are the
“real value”, we can compute the percentage of this value among the sum of all six movers status changes at the seam (December to January). This will give us the percentage of spurious transitions. As previously described the November to December transition is considered the most precise report and also the best reference point to compare the seam transition because closest less prone to seasonal effects in the job market. Table 4.11 provides these calculations for two seam points that will be object of further analysis.

Table 4.11
Percentage of Movers Spurious Transitions at the Seam When Compared to the November to December Transitions.

<table>
<thead>
<tr>
<th>Sum of six movers transitions status</th>
<th>Nov to Dec (%)</th>
<th>Seam (Dec to Jan) (%)</th>
<th>% of spurious transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CQ seam 1998_99</td>
<td>0.80</td>
<td>8.82</td>
<td>90.93</td>
</tr>
<tr>
<td>EHC seam 2002_2003</td>
<td>0.65</td>
<td>9.18</td>
<td>92.92</td>
</tr>
</tbody>
</table>

The table shows that roughly 90% of those that moved at the seam are spurious (an overestimation of the phenomena). Table 4.11 gives reasoning to the initial idea of using the movers transitions at the seam as a good approximation for seam bias. In other words, 90% is bias and 10% are real transitions. We now have a dependent variable that can be used in multivariate analysis in order to study the relationship between bias and bias factors at a person or design level. Another way to look at the issue is to think about a baseline of error common to each seam point.

In order to find out what variables are affecting the error at the seam two exploratory logistic regressions are performed on seam 4 (CQ1998-99, two year reference period) and on seam 6 (EHC 2002-03, two year reference period). Logistic regression appears to be the best fit for the kind of data in hand. In comparison to other techniques, such as discriminant function, logistic regression does not have assumptions about the distributions of the predictor variables, the predictors do not have to be normally distrib-

\[16\] For the first case in example the % of spurious transitions is \(= 100 - (0.80 \times 100/8.82) = 90.93\)
uted, linearly related, or of equal variance within each group (Tabachnick & Fidell, 2001 p. 517).

The dependent variable will be the group membership 1 (seam bias) or 0 (no bias). In order to compute group membership each answer will be classified at each seam according to the following rules: people reporting a transition or *movers* (EU, EN, UE, UN, NE, NU) will be classified as belonging to the seam effect group (coded as 1). People not reporting a transition or *stayers* (EE, UU, NN) will belong to the no change group (coded as 0). Predictor variables, their categories and measurement level are delineated in Table 4.12. Further description of each variable follows.
Table 4.12
Predictor Variables Used in the Logistic Regression, their Categories and Measurement Level

<table>
<thead>
<tr>
<th>Variable</th>
<th>Categories</th>
<th>Measurement level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Person level characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-proxy answer</td>
<td>Self-self</td>
<td>Categorical, dummy coded, 2 vectors (see table 4.12)</td>
</tr>
<tr>
<td></td>
<td>Proxy-proxy</td>
<td>Reference category (self_self)</td>
</tr>
<tr>
<td></td>
<td>Self-proxy &amp; proxy self</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>In years</td>
<td>Continuous</td>
</tr>
<tr>
<td>Years of completed education</td>
<td>In years</td>
<td>Continuous</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>Categorical, dummy coded</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>Reference category: Male (1)</td>
</tr>
<tr>
<td>Race</td>
<td>White</td>
<td>Categorical, dummy coded</td>
</tr>
<tr>
<td></td>
<td>Not white</td>
<td>Reference category: White (1)</td>
</tr>
<tr>
<td>Complex labor force history</td>
<td>Number of status changes within the previous year</td>
<td>Continuous</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Design level characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Having the same interviewer at both waves</td>
<td>Yes</td>
<td>Categorical, dummy coded</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Reference category: Yes (1)</td>
</tr>
<tr>
<td>Interview lag from December of first wave</td>
<td>Day elapsed from December of time $T−1$ to the interview of the first wave</td>
<td>Continuous in number of days</td>
</tr>
<tr>
<td>Interview lag from January of the second wave</td>
<td>Day elapsed from January of time $T−2$ to the interview of the second wave</td>
<td>Continuous in number of days</td>
</tr>
</tbody>
</table>

The selection of the above variables is dictated by the knowledge gained from the previously reviewed studies on factors affecting the magnitude of seam effect plus some other hypothesis matured during the course of this investigation.
4.4.1 *Person level characteristics predictor variables*

*Self-proxy answer.* The studies done on the self-proxy issue at the seam (Moore & Kasprzyk, 1984; Murray, Michaud, Egan, & Lemaître, 1991; Vick & Weidman, 1989) report contrasting results, the increase or decrease of status changes at the seam depends on the kind of variables used. Because approximately 40% of answers in the PSID are proxy proxy (including the combinations proxy self- self-proxy) as shown in Figure 4.6, the variable self-proxy appears to be an important predictor to be included in the exploratory models. The reader should keep in mind that in our case the proxy is either the Head or the Wife, and that receiving the status of Wife is dependent for either being legally married or from cohabitating for at least one year. In order to include in the model the variable self-proxy answer (categorical) a transformation is necessary. A categorical variable is coded into N-1 number of levels dummy vectors (Pedhazur, 1997). These new vectors need to be independent (orthogonal) of each other. In this case, because self-proxy and proxy-self were combined due to the small sample size, we have 3-1=2 new dummy variables as showed in Table 4.13. The group object of comparison should be the group with the highest frequency i.e. Self-Self (00).

![Table 4.13](image)

<table>
<thead>
<tr>
<th>Group</th>
<th>SPPS (SS)</th>
<th>PP (SS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Self</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Self-proxy &amp; proxy-self</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Proxy-proxy</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

The very low correlation of error between self and proxy answers investigate at the beginning of this chapter (table 4.4) justifies accepting the logistic regression assumption of independence of errors (responses of different cases are independent of each other).
Age (AGE). Age was found to amplify inconsistencies of reports at the seam by Hill (1987) and Jäckle and Lynn (2004) thus creating higher seam bias. Age in the PSID is measured in number of years. Respondents become eligible to be interview when they form an economically independent household. The range of age goes from 15 years old to 99 with a mean around 43 years old.

Years of completed education (EDUC). This variable was included as a surrogate for cognitive capacity in order to test if that has an effect on the seam bias. The variable is measured using the completed grade in school where 17 is the maximum and indicates at least some post-graduate work.

Sex (SEX). Although sex was not found to be a predictor of seam bias in the previous studies, it is considered important to be included in the model as a control variable. Male is the reference category coded as 1.

Race (RACE). Race was found to have an effect at the seam by Hill (1987). African Americans were found to be more likely to provide higher inconsistencies at the seam than Whites, controlling for education, sex, age and income. Race is dummy coded in the analysis where White is the reference category coded as 1.

Complex labor force history (COMPLIC). This variable is hypothesized to have a strong impact on the model. The hypothesis is that at an increase of job changes there is an increased likelihood to report a spurious status changes at the seam. People with complicated job histories, in and out of the job market, have more chances to make more errors that somebody who hold a steady job for the entire reference period. Constant wave responding and telescoping, for example, can create spurious status changes. Hill (1987) found a statistically significant negative relationship between number of months with the current employer and amplifying inconsistencies at the seam. Complex labor force history is measured counting the number of status change during the year before the seam (Time $T-1$). In this case people reporting multiple changes within a month (the previously mentioned unclassifiable cases) are included in the model because it is a strong indication of complexity of job status. People with multiple status changes within the wave have also more chances to have real status changes at the seam.
4.4.2 Design level characteristics predictor variables

Having the same interviewer (SAMEINT). Wick and Weidman (1989) found that having the same interviewer at both waves decreased seam effect in the SIPP panel. The SIPP was a face to face data collection with interviews approximately every four months. In the PSID the interviews are by phone and now every two years. For this reason, the initial hypothesis that having the same interviewer could potentially reduce seam effect appears difficult to be verified but it is worth exploring the possibility of it. The variable is measured using interviewers ID where in the case of match a value of 1 is assigned. Unfortunately in 2005 the interviewers’ ID numbering system changed, thus rendering the computation impracticable for the EHC seam.

Interview lag from December of first wave (ELAPS_D_1ST_INT). Because PSID interviews start in February and end in November (see chapter 3), it is plausible to think that people interviewed in February might have better memory regarding December of time $T-1$ than people interviewed in November. This variable is measured counting the number of elapsed days from the first interview to December 31st of time $T-1$ of the first wave.

Interview lag from January of second wave (ELAPS_D_2ND_INT). Same reasoning as above but just for the memory of January of $T-2$ of the second interview. The variable is measured counting the number of elapsed days from the second interview to January 1st of time $T-2$ of the second wave.

4.4.3 Results from CQ and EHC logistic regressions predicting seam bias

Two logistic regressions were fitted on the following seam points: CQ 1998_99 two year reference period, and seam number 6, EHC 2002_03 two year reference period. These two seam points are chosen because focus of this dissertation in analyzing CQ and EHC differences. All the predictor variables hypothesized to have an effect on the model were
entered simultaneously (direct logistic regression). The analysis was performed with SPSS for Windows version 14. Table 4.14a, 4.14b report the results from the analyses.

Table 4.14a  
CQ 1998_99 Logistic Regression Output

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Odds Ratio</th>
<th>Sig.</th>
<th>95% CI for Odds Ratio</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPPS (SS)</td>
<td>1.086</td>
<td>.664</td>
<td>.748 - 1.578</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP (SS)</td>
<td>.806</td>
<td>.024</td>
<td>.669 - 0.972</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>.983</td>
<td>.000</td>
<td>.977 - 0.989</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDUC</td>
<td>.926</td>
<td>.000</td>
<td>.895 - 0.959</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEX (Male =1)</td>
<td>.725</td>
<td>.000</td>
<td>.613 - 0.858</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RACE (White =1)</td>
<td>.796</td>
<td>.010</td>
<td>.669 - 0.947</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPLIC</td>
<td>1.827</td>
<td>.000</td>
<td>1.663 - 2.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAMEINT</td>
<td>.920</td>
<td>.484</td>
<td>.728 - 1.162</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELAPS_D_1ST_INT</td>
<td>1.001</td>
<td>.265</td>
<td>.999 - 1.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELAPS_D_2ND_INT</td>
<td>.999</td>
<td>.086</td>
<td>.997 - 1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2.121</td>
<td>.366</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Nagelkerke $R^2 = .080$
N = 8,019

Table 4.14b  
EHC 2002_03 Logistic Regression Output

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Odds Ratio</th>
<th>Sig.</th>
<th>95% CI for Odds Ratio</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPPS (SS)</td>
<td>1.113</td>
<td>.573</td>
<td>.767 - 1.615</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP (SS)</td>
<td>.945</td>
<td>.546</td>
<td>.787 - 1.135</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>.984</td>
<td>.000</td>
<td>.978 - 0.990</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDUC</td>
<td>.906</td>
<td>.000</td>
<td>.874 - 0.938</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEX (Male=1)</td>
<td>.692</td>
<td>.000</td>
<td>.586 - 0.817</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RACE (White=1)</td>
<td>.681</td>
<td>.000</td>
<td>.578 - 0.803</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPLIC</td>
<td>1.847</td>
<td>.000</td>
<td>1.718 - 1.985</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELAPS_D_1ST_INT</td>
<td>1.001</td>
<td>.532</td>
<td>.999 - 1.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELAPS_D_2ND_INT</td>
<td>1.001</td>
<td>.108</td>
<td>1.000 - 1.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2.25</td>
<td>.073</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Nagelkerke $R^2 = .121$
N = 8,212
After an inspection of the two logistic regressions results some communalities appear: the self-proxy answers have an effect in the CQ model but not in the EHC model. When there is a combination of proxy_proxy in comparison to a self_self answer there is a reduction of seam bias but only in the CQ interview. The second main finding is that the design level characteristic variables do not have an effect in improving the fit of the model. The time lag either from the first or the second interview is maybe too long anyway in order to have an impact on the seam bias. Having the same interviewer does make any difference either. This might be the case that after two or more years between the two waves, the retrieval cues generated by keeping the same interviewer are forgotten.

The two models provide very close results for the predictors. The CQ model accounts for about 8% of the variance while the EHC for about 12%. Although these results are not impressive, the reader should be reminded that the dependent variable contains some error because it was not possible to distinguish between real transitions and spurious transitions.

An increase in age slightly decreases the odds of having seam bias. Perhaps older people have been in the panel longer and for this reason there might be more motivated (optimizers) in trying harder to remember their job history thus decreasing the bias at the seam.

Education can be seen as a surrogate measure of cognitive capacity and it goes in the expected direction, people with higher education have lower seam bias. Higher education can also mean having more stable jobs.

Females are 28% (CQ) or 31% (EHC) more likely to show seam bias than Males. Non-Whites are 21% (CQ) or 32% (EHC) more likely to show seam bias than Whites. Because the model is controlled for the number of labor force transitions, age, and education, these results are not easy to explain. More research is needed on this regard, possibly employing datasets with validation data.

People with more status changes (COMPLIC) within the wave, are more likely to show bias at the seam. Because complicated life history is measured counting the number of status changes within the wave, we can interpret the odds ratio of 1.8 for both CQ and
EHC to signify that an increase of 1 on the complicated life history measure increases the odds of showing seam bias that roughly two times. COMPLIC is the variable with the highest odds ratios. The more changes in job status during the year, the more chances to make a mistake and report a transition at the seam.

Lastly, for the self-proxy variable we have mixed results. In the CQ the combination proxy proxy (self_self as reference) had an odds ratio statistically significant and lower than 1, meaning that self-self answers show 20% more seam bias when compared to proxy_proxy answers. An answer that is proxy at the first wave and at the second wave is more likely to be a simplification of the job history of the partner, thus resulting in fewer chances to be inconsistent at the seam. The effect however disappears in the EHC seam. In reminding that the correlation of error between self and proxy answers was lower in the EHC condition (Table 4.4) a possible explanation could that in the EHC the respondent has more retrieval cues for the proxy report thus being as likely to make mistakes as a self answer at the seam. The self-proxy issue at the seam does not inform about the quality of self and proxy report within the wave. A study with validation data can answer this question.
Summary and Conclusions
Summary of the results

This dissertation is written from the point of view of measuring a particular error in survey research, specifically an error peculiar to longitudinal surveys: the seam effect. Seam effects were first found in the PSID for variables such as unemployment compensations and food stamp recipiency (D. H. Hill, 1987). The present study adds to the literature of seam effects for the PSID.

All existing research has examined seam effects with data collected through a standardized conventional questionnaire. This study investigates the magnitude of seam effects in labor force data in the PSID focusing on a comparison between CQ data collection and the event history calendar (EHC). Based on previous studies showing higher data quality of EHC in terms of reduction of underreporting and increased precision of the location of events in time when compared to CQ reports (Belli, Shay, & Stafford, 2001; Belli, Smith, Andreski, & Agrawal, 2006) the hypothesis of this dissertation was to expect a reduction of seam effect when the PSID introduced the EHC data collection method. The analysis supports this hypothesis showing that the EHC seam effect is of less magnitude than the CQ seam effect. The EHC also reduced the correlation of error between self and proxy answers at the seam in comparison to CQ. Carefulness in interpreting these results is required because differences in question order and wording were not controlled. Consequently the non experimental nature of the data precluded a clean comparison of the CQ to EHC seam effects.

The analysis brought to light a previously undocumented phenomenon, the “within-wave seam effect”, found when the PSID moved to a two year reference period in the CQ data collection mode. The explanation for the within-wave seam effect lies in the design of the CQ questionnaire. Questions about labor force status for time $T-1$ were asked at the beginning of the interview and those regarding $T-2$ took place 30 minutes later. They key is that $T-2$ questions were not as detailed as $T-1$ and for this reason did not provide enough retrieval cues to the respondent. The within-wave seam exemplifies how questionnaire design can create seam effects during the same data collection period.
Supporting this idea, when the questionnaire was changed in the EHC waves and the three types of labor force questions were asked for each of the 24 months at the same time, the within-wave seam disappeared.

Status changes from “not in the labor force” to employment and from employment to “not in the labor force” status (NE and EN) were found to manifest the highest seam bias. Difficulty in conceptualizing the not in the labor force definitions and in remembering a non event (the “not in the labor force status” is assigned when respondents report not having a job and not being actively looking for one) were advocated as possible sources of bias.

The magnitude of the one year recall period was always lower than any of the two year recall period seam bias. That means that EHC was not able to offset the increase of bias due to the increase of the length of the recall period. Even if EHC gives more retrieval cues than CQ, remembering the job status occurred more than two year from the time of the interview is a difficult task.

Lastly, the driving factor predicting seam bias appears to be having a complicated job history, operationalized as the number of labor force status changes within the reference period at time $T -1$. People changing labor force status frequently have more chances to make mistakes at the seam than people with a stable job history. However, due to the absence of validation data, we could not precisely describe nor explain who are the people and what kind of variables specifically predict being subject of error at the seam, since it was not possible to disentangle real changes from spurious changes at the seam.

PSID data users should be aware of possible seam effects in the variables they analyze and be careful when making interpretations of status changes at the seam between two waves. The magnitude of seam effect with data collected with EHC needs at least another wave of data collection to be better understood. In fact, it is plausible that with more practice and better interviewer training, seam effect could be potentially further decreased with EHC.
Possible solutions to further reduce seam effects in the PSID

Seam effects for labor force status changes in the PSID is still present even after the switch in data collection to EHC. We have to accept that the fact “that event history calendars show mostly encouraging gains in data quality in comparison to standardized interviewing indicates that it is not a panacea that will cure all ills associated with forgetting” (Belli & Callegaro, in press-b). To my awareness, the only other seam effect study with data collected with EHC is described in Carroll (2006). Although there is no comparison with other seam points because the analysis is done using the first two waves of the panel, the author found seam bias with magnitude five times larger than the within wave changes for inflow (entering the job market) and outflow (leaving the job market). HILDA used EHC for labor force history with a reference period of one year. The just mentioned overlapping between waves 1 and 2 enabled the author to compute percentages of consistent answers. The results show that “27% of people gave answers that were never consistent (where the categories are unemployment and not employment), 40% of people were consistent the entire time and the remaining were consistent in between one to eight of the 9 periods examined” (Carroll, 2006, p. 301). HILDA seam effect results point to the same direction of the results from this dissertation; EHC data collection by itself is not effective enough to dramatically reduce seam effect. Some extra strategy could be applied to further reduce seam effect, for example the use of dependent interviewing.

When looking at the literature about strategies successful in reducing seam effects, dependent interviewing appears the most successful (Jäckle, in press). PSID does not use any form of dependent interviewing. One of the main reasons is that the PSID interviews one person only per household and not both Head and Wife. Thus in a household where Head and Wife are both present half of the labor force answers for one of the partners will be proxy answers. Proxy answers are not a problem per se in using dependent interviewing. What is really problematic is when there is a switch from wave to wave between self and proxy or vice versa, or when there is a new Head or a new Wife. In that
case question wording and ethical issues arise. When information is collected from one respondent and given to another respondent, question wording can become awkward. In order to avoid complicated\textsuperscript{17} verbiage such as “he/she told us”, “your wife/husband reported…” one strategy is to use sentences such as: “according to our records, when we last interviewed you…” to solve the problem. Moreover, ethical issues and confidentiality of the answers are at stake when dependent interviewing is used to feed information to a respondent different from the previous wave. One solution adopted by the U.S. Census, for example, is to ask permission to disclose the respondent’s answers to other household members\textsuperscript{18} (Chan & Moore, 2006). From the analysis of Figure 4.5 it appears that there is only roughly less than 5% of self/proxy, proxy/self switch between waves, thus rendering the problematic issue of feeding information to another respondent applicable to a very small part of the sample. What is really challenging in using dependent interviewing is the necessary extra programming time, testing and, consequently, overall increased cost (Hoogendoorn, 2004). What requires extra effort and testing is to make sure that the right information is given to the right respondent. The benefits obtained are a reduction of the bias at the seam, and also an increase of the overall quality of the within wave reports (Jäckle, in press; Moore, Bates, Pascale, & Okon, in press).

What would be worth experimenting for the PSID is the use of dependent interviewing in conjunction with EHC, as in the Italian Longitudinal Household Survey (Schizzerotto, 2002), for example. It would be interesting to do another PSID validation study comparing data collected in three ways: one that merely uses EHC, a second one using conventional questionnaire dependent interviewing, and a third one where dependent interviewing and EHC are combined. With this design it would be possible to identify which variables benefit from each data collection method, in order to use them in a tailored fashion on a full scale of the panel. For example, from the present study it appears

\textsuperscript{17} Complicated also from a software programming point of view.

\textsuperscript{18} For example the SIPP uses the following formulation: “One last thing... We re-contact households once over a 4- month period to update information. If we talk to someone else in your household next time, instead of you, is it OK if we use your answers as a starting point?” (Chan & Moore, 2006, p. 22)
that something needs to be done in order to provide respondents with more retrieval cues and maybe a better measurement instrument to identify periods of “not in the labor force”.

**Study limitations**

One of the limitations of this study is the unavailability (at the time of this writing) of the 2005 wave weights. This precluded the presentation of the results with weighted data for all the waves. An analysis with weighted data for the 27th and 28th waves (not shown) did not illustrate any major change in the overall pattern. Attrition could potentially influence the magnitude of the seam effect and weights can compensate for that. It is however true that attrition in the PSID is extremely low. Low attrition is likely to be the explanation of the negligible differences between weighted and unweighted data in the above described comparison between weighted and unweighted data of the 27th CQ and 28th CQ waves.

The absence of validation data precluded a more precise analysis of the predictors of seam bias in the logistic regression analyses presented in section 4.4. With validation it would be possible to have a variable that for each subject tells if the transition at the seam is “real” or if it is seam bias. In other words, it would be possible to disentangle the respondents who legitimately changed their labor force status at the seam from the respondents who did not, enabling to study the characteristics of the latter. However, good validation data on labor force status are extremely difficult to obtain from external sources. Although test-retest approaches (asking the same questions at different points in time regarding the same reference period) have been used when external data are unavailable, these approaches suffer from reliability issues since even asking the same questions in the same interview produces some unreliable answers.

The non experimental nature of the data makes difficult to isolate each element that contributed to the increase or decrease of seam bias. Randomizing respondents to either CQ or EHC keeping the questionnaire as close as possible for two waves would have been a solution to study the net effect of each data collection method.
Only one CQ two year recall period seam point is available and at the present time we have only one EHC two year recall period seam point as well. As shown in Figure 4.7 there are fluctuations of seam bias for the one year recall period. Starting with the availability of the data from the 2007 wave we will have more than one EHC seam point that can be analyzed, enabling to study the stability of the seam bias over the course of several years.

Another limitation of the present study is that at the present time pre-release data have been used to compute the EHC two year recall period seam bias. Official release data might change the shape of the distribution although PSID staff has provided me with assurances that on the variables used in the computation of the labor force status changes not much data cleaning and editing is actually done.

Lastly, we do not know if the magnitude of the bias at the seam for labor force status variable is higher or lower than for other variables in the PSID. To my knowledge, no other study regarding seam effects for the periods under investigation has been published. Only with additional research will it be possible to assess how EHC was effective in reducing the magnitude of the seam for different kind of variables.

**Conclusions and perspectives of future research**

Seam effects are common across different types of variables and have been found in different panels across the world. The common denominator of the methods that thus far have been proven effective in reducing seam effects is that they use different memory aids in helping the respondent retrieve information.

Dependent interviewing, calendar aided interviewing, and having the same interviewer all provides respondents with more retrieval cues than with a conventional questionnaire. With dependent interviewing the respondents are reminded of the previous wave’s answer and/or inconsistencies are pointed out thus inviting them to try harder in retrieving the facts and providing the answers. The other explanation is that respondents try to be consistent with the final effect of reducing seam bias. The calendar method used
by Kominski (1990) visualizes the previous interview and works as a sort of diary providing retrieval cues similar to dependent interviewing but also with the location of the previous answers on a timeline. If the same interviewer administers the questionnaire in each wave, then the same interviewing setting is recreated, or, in other words, Tulving’s memory encoding specificity principle is used. Grouping the questions by reference period and not by item (Rips, Conrad & Fricker, 2003) reduces satisficing behaviors and “forces” the respondent to think more about the events.

This dissertation contributes to the literature suggesting the event history calendar as another method to reduce seam effect. It also contradicts the findings by Vick and Veidman (1989) that keeping the same interviewer at both waves reduces seam effect probably because in the PSID the distance between the two interviews is of two years and not of four months. Lastly, it shows how it is possible to obtain seam effects within the same interview due to questionnaire design.

It should be noted that seam effects have been studied on a small range of variables. We need more research on other variables to understand which ones are more sensitive to seam effect and which ones are less sensitive or do not present seam bias. Without validation data it is impossible to judge whether an increase of status change at the seam transition point is due completely to a seam bias or that the number of “real” status changes is higher at the seam. Validation studies will also guide researchers who develop statistical adjustments for seam effects because they provide information on the distribution properties.
References


PSID staff. (2005a). *PSID Question by question objectives*. Ann Arbor, MI.


Appendix.

Extended questionnaires flowcharts
SECTION B: EMPLOYMENT OF THE HEAD

B1 We would like to know what you do--are you (head)?

1. Working now

2. Only temp laid off, sick leave or maternity

3. Looking for work, unemployed

4. Retired

5. Disabled

6. Keep house

7. Student

8. Other

B3 Are you (head) doing any work for money now at all?

1. Yes

2. No

B4 (Are you (head) self-employed, employed by somebody else or both to B11 What is the name of your (head's) employer?

B39 In which months during (previous year) were you working for (name of employer from B11) as your main job?

B40 Did you have any other main job-employer at any time during (previous year)

1. Yes

2. No

B41 to B42 In which months during (previous year) were you working for (name of employer from B41c)

B59 Did you have any other main job-employer at any time during (previous year)

1. Yes

2. No

Pink work hist. suppl.

Section C Head is not working now

B60 to B72
SECTION C: HEAD IS NOT WORKING NOW (No to B3)

C1 Have you been looking for work during the past 4 weeks?
  1. Yes
  2. No

C2 to C4 Have you (head) ever done any work for money?
  1. Yes
  2. No

C5 In what month and year did you last work?
  199x
  199x+1
  DK which
  Other year
  Before 199x
  DK year at all

C6 Were there any times in (previous year) when you were looking for work?
  1. Yes
  2. No

SECTION D

C31 In which months during (previous year) were you working for (name of employer from C14a)?

C32 Did you have any (other) main job-employer at any time during (previous year)
  1. Yes
  2. No

C34 In which months during (previous year) were you working for (name of employer from C33c)?

C52
C52 to C64. Did you miss any work in (previous year) because you were unemployed and looking for work or temporarily laid off?

1. Yes
   - C66a In which months were you unemployed for at least one week?

2. No

C67 Were there any weeks in (previous year) when you didn't have a job and were not looking for one?

1. Yes
   - C69a Which months had at least one week when you didn't have a job and were not looking for one?

2. No

C74 (Besides the job(s) you have already told me about,) did you (head) have an extra job or other way of making money in (previous year)?

1. Yes
   - C83 In which months during (previous year) were you working (for that employer/on this extra job)?

2. No

C86 Did you have any extra job in (previous year)?

1. Yes
   - C95 In which months during (previous year) were you working (for that employer/on this extra job)?

2. No

C98 Did you have any extra jobs in (previous year)

1. Yes

2. No

SECTION D

Ask C87-C98 Again
Flowchart 2 CQ T–I Head Section

1999 and 2001 T-2 reference period

R25/R32 Did you (head of wife) work at a job of business at any time during (1999)?

Yes

R28/R35 During which months was that?

No

R30/R37 Were there any months in (1999) in which you were unemployed and looking for work for at least one week?

Yes

UNEMPLOYMENT

Which months were those?

No

EMPLOYMENT
SECTION BC: EMPLOYMENT OF THE HEAD

BC1 We would like to know what you do—are you (head)?

1. Working now
2. Only temp laid off, sick leave or maternity

BC2 In what year did you retire?

4. Retired
3. Looking for work, unemployed
5. Disabled
6. Keep. house
7. Student
8. Other

GO TO BC6

BC3 Have you done any work for money since January 1, (previous 2 years)

1. Yes
2. No

GO TO BC7

BC4 I'd like to know all about the work for money that (you/HEAD) have done for the past two years, from January 1, (previous 2 years)

FOR EACH EMPLOYER BC6 When did you start and when did you stop working for this employer? Please give me all of your start and stop dates if gone for work more than once

BC6 When did you start and when did you stop working for this employer?

BC7 Now I'd like to know about times during (2 previous years) in which (you were/he was) not working at all for pay. First, I'd like to know about any times when (you were/he was) NOT looking for work, and next I'd like to know about any times when (you WERE/he WAS) looking for work. Was there any time in (T-1 and T-2) when you did not have a job and were not looking for work?

1. Yes
5. No

When was that?

BC8 Was there any time last year, (in previous year), when (you were/he was) unemployed and looking for work?
BC9 Was there any time in the year before last, in (previous 2 years), when (you were/he was) unemployed and looking for work?

1. Yes
5. No

When was that?

Question BC15 to BC60 referred to the previous jobs (kind of job, pay...)

For people who did not work in the previous 2 years

BC62 Have you ever done any work for money?

5. No

1. Yes

BC63 In what month and year did you last work?

For everybody

BC64 (Have you/Has HEAD) been looking for (another job/work) during the past four weeks?

1. Yes → GO TO BC65-BC67

5. No → GO TO SECTION DE