THE SURVEY OF INCOME AND
PROGRAM PARTICIPATION

DEPENDENT AND INDEPENDENT DATA COLLECTION IN PANEL SURVEYS: ANALYSIS OF 1985, 1986 SIPP OCCUPATION AND INDUSTRY DATA

No. 166

D. H. Hill
Survey Research Institute/
University of Toledo

U. S. Department of Commerce BUREAU OF THE CENSUS

Dependent and Independent
Data Collection in Panel Surveys:
Analysis of 1985-1986 SIPP Occupation and Industry Data

Daniel H. Hill
Survey Research Institute
University of Toledo
June 30, 1992

Note: This research was sponsored by a Joint Statistical Agreement (JSA 91-17) between the U.S. Census Bureau and the University of Toledo. The findings, recommendations, or conclusions are not endorsed by the Government, and do not necessarily reflect policies of the Census Bureau. The author would like to thank Dan Kasprzyk for suggesting this research topic. Rajendra Singh and Tom Scopp of the Census Bureau provided a number of helpful comments on earlier versions of this paper and Jim Lepkowski, Steve Pennel and David Miller of the University of Michigan assisted in obtaining the data. Finally, the author would like to thank Mary Louise Mikula of the Survey Research Institute for her many hours of coding and recoding the occupation characteristics data. Any errors are the author's responsibility.

TABLE OF CONTENTS

Dependent and Independent
Data Collection in Panel Surveys:
Analysis of 1985-1986 SIPP Occupation and Industry Data

Daniel H. Hill
Survey Research Institute
University of Toledo
June 30, 1992

I.1 Background

The 1985 and 1986 SIPP Panels present a rare opportunity to study the effects of data collection mode in panel surveys on the stability of two important survey measures--occupation and industry. The reason is that the 1986 panel used a question to determine whether occupation and industry questions should be asked in the next wave (i.e., dependent data collection mode), whereas the 1985 panel did not (i.e., independent data collection mode). If the respondent said no to this question then occupation and industry codes from the prior waves were carried forward to the current wave. Since these panels overlap in historical time and each is a probability sample of the same population (and since there is no evidence of panel conditioning), the combined data represent a randomized experiment with the treatment being the use of the screening question. The outcome measures consist of the month-to--month gross change in occupation and industry of employment.

Occupation and, to a slightly lesser extent, industry are notoriously unreliable measures. Noise enters the measurement process via variation in how the respondent describes his/her duties, in how these descriptions are recorded by the interviewer and in how the coders interpret it.[1] Thus, even in the absence of real change in occupation and industry, independent data collection will show greater instability than will dependent simply because more coding is done.

---

[1]From 1968-1986 the Panel Study of Income Dynamics routinely conducted a 100% check-coding operation for occupation and industry. The results were that, with experienced coders, the reliability of coding the same responses was approximately 92% for occupation and 94% for industry.

Just as there are _a priori_ reasons to expect the Independent measurement method to overstate change in occupation and industry, there are reasons to expect the dependent method to _understate_ true change. Industry and occupation questions are asked and coded for those with the same employer if and only if the respondent answers yes to the following 'screening' question:

" **2b. Have ...'s main activities or duties for this employer changed during the past 8 months?** "

The propensity of respondents to fail to recall or report such retrospective changes is both substantial and well documented (see e.g. Mathiowetz, 1986, or Neter and Waksberg, 1965). Furthermore, since the response task is considerably reduced, there is an incentive for respondents to say 'no' when asked item 2b even when there has been some change.

The extent to which dependent data collection methods understate true occupational change is an important question. Unfortunately, without reliable validating information, very little can be said about it. There is some external evidence that the amount of true change missed by dependent collection methods might be considerable. In the CPS Job Mobility Study (U.S. Department of Commerce, 1975), for instance, 'expert' coders conducted detailed analyses of observed occupation changes from independent mode data and classified them into those due to coding errors, ambiguous responses on the part of respondents and true change. While most of the observed 'change' was found to be due to either coding error or ambiguous wording, it was concluded that the vast majority of 'true' occupation changes would be missed by dependent data collection methods. This conclusion seems to be rather damning for dependent collection methods until it is realized that any change in how the respondent described his duties, _or the order in which the respondent mentioned various duties_, was sufficient for the observed occupational change to be classified a 'true' change. Since the respondent is not told that the _order_ in which he lists his duties matters, it would not be surprising if much of the 'true' change observed in the Job Mobility Study was due simply to variance in salience of specific duties on the part of respondents from one interview to the next.

I.2 Approach and Organization

To assess the effects of collection mode on the observed stability of occupation and industry, we first merged information from the 1985 and 1986 SIPP Full Panel Longitudinal Research Files for the calendar months January 1986 through April 1987. The primary sample employed in this report consists of all 100-level individuals from the 1985 and 1986 Panels who were 1) sixteen years of age or older in January 1986; 2) were interviewed in each of the four waves covering the January 1986-April 1987 period; and 3) had an employer for at least two months during the period.

Three distinct methods of measuring and analyzing industry and/or occupation change are employed. In Section II, we examine changes in main-job (WS1) occupation and industry, and their associations with changes in work hours, wages and employers without regard to its source. Main-jobs which change to second jobs are considered changes just as main-jobs which are abandoned and replaced with other main-jobs. In Section III, we adopt a more restrictive definition of change by not counting changes to second jobs and by requiring a three month absence from an occupation for it to be considered to have ended. In this section, our major focus is on the identification of those occupations and industries for which method of collection is most important. Finally, in Section IV, we confine our attention to the first 'non-left-censored' occupational spell using the more restrictive definitions of change developed in Section III. Our focus in this section is on the implications of collection mode on the estimated parameters of event history models.

II Main-Job Occupation and Industry Changes and Their Association with Change in Hours, Wages and Employers

The first order of empirical business is to assess the extent to which the overall amount of observed occupational and industrial change is affected by collection method, and how these changes relate to contemporaneous changes in other job characteristics.

The simplest way of doing this is to concentrate on the main-job (WS1) during the January 1986 through April 1987 period and create a record for each occurrence of a change from one month to the next in the occupation or industry code. The distribution of changes in hours, wages and employer identifiers between the corresponding months can then be calculated. In doing so we will confine our attention to intra-marginal changes by excluding occupation and industry changes for those entering or leaving the labor force.

Table II.1 presents the distribution of changes in hours for all person-month changes in occupation by mode of collection. The bottom row of the table presents the total number of occupational changes estimated for the population during the 15-month period. This total is nearly five-times as large for the independent-mode data (approximately 123 million changes) as for the dependent data (approximately 26 million changes). The table also reveals that the majority of occupation changes in the independent-mode data had no corresponding change in work hours. With the dependent-mode data, on the other hand, more than half of the occupation changes were accompanied by an increase or decrease of 10% or more in work hours and an additional seven percent had changes ranging from one to ten percent in absolute value.

Since this same general pattern appears with respect to wage change of hourly workers (Tables II.2 and II.5) and employer changes (Tables II.3 and II.6) for both occupation and industry, it seems safe to conclude that, as in the CPS Job Mobility Study, most of the observed 'change' with independent data collection methods is noise. The distribution of changes in employer for industry changes (Table II.6) provides particularly strong support for this conclusion. With independent mode data, nearly three quarters (73.5%) of the month-to-month industry 'changes' were for individuals working for the same employer in both months. While it is possible for individual firms to change industry, such changes are quite rare. Somewhat less rare are transfers from one subsidiary of a conglomerate to another which could entail an industry change without an employer change. However, these changes are also likely to be uncommon. Thus, the distribution for the dependent-mode data, in which more than 80% of the industry changes are accompanied by an employer change, is far more plausible than the independent-mode distribution.

These distributions of employer changes for industry 'changers' also provide some moderately convincing evidence that the dependent mode data captures most of the true industry change in the population. Specifically, the dependent-mode estimate of the total number of concurrent changes in employers and industries (17 million) is less than eight percent lower than the corresponding total (18.35 million) observed using the independent collection method. If all 'real' industry changes involve an employer change, and if the independent-mode data captures all real changes, then these data suggest that the dependent-collection methodology also captures most (more than 92%) of them.

Table II.1
Distribution of Changes in Work Hours
All Month-to-Month Main-Job Occupation Changes
(All Workers: February 1986 - April 1987)

Percent Change in Hours	Independent Mode (1985 Panel)	Dependent Mode (1986 Panel)
<-24% (Decrease)	10,746 8.7%	5,713 21.8%
-24% to -10%	10,466 8.5%	2,514 9.6%
- 9% to -1%	4,940 4.1%	1,133 4.3%
0% (No Change)	72,175 58.8%	10,302 39.3%
1% to 9% (Increase)	5,235 4.2%	816 3.1%
10% to 24%	10,344 8.4%	2,048 7.8%
> 25%	8,784 7.1%	3,667 14.0%
Total	122,663 100.0%	26,188 100.0%

Amounts in thousands.

Table II.2
Distribution of Wage Changes
All Month-to-Month Main-Job Occupation Changes
(Hourly Workers: February 1986 - April 1987)

Percent Change in Wages	Independent Mode (1985 Panel)	Dependent Mode (1986 Panel)
<-24% (Decrease)	5,052 7.2%	2,670 17.1%
-24% to -10%	5,447 7.8%	2,043 13.0%
- 9% to -1%	13,571 19.2%	2,184 13.9%
0% (No Change)	34,239 48.6%	4,092 26.1%
1% to 9% (Increase)	6,597 9.3%	1,710 10.9%
10% to 24%	3,129 4.5%	1,643 10.5%
> 25%	2,454 3.5%	1,326 8.5%
Total	70,491 100%	15,671 100%

Amounts in thousands.

Table II.3
Distribution of Employer Changes
All Month-to-Month Main-Job Occupation Changes
(All Workers: February 1986 - April 1987)

	Independent Mode (1985 Panel)	Dependent Mode (1986 Panel)
No Change in Employer	104,606 85.3%	9,121 34.8%
Change in Employer	18,057 14.7%	17,066 65.2%
Total	122,663 100%	26,188 100%

Amounts in thousands.

Table II.4
Distribution of Changes in Work Hours
All Month-to-Month Main-Job Industry Changes
(All Workers: February 1986 - April 1987)

Percent Change in Hours	Independent Mode (1985 Panel)	Dependent Mode (1986 Panel)
<-24% (Decrease)	7,194 10.4%	5,385 26.2%
-24% to -10%	5,641 8.2%	2,000 9.7%
- 9% to -1%	2,704 3.9%	708 3.4%
0% (No Change)	39,427 57.0%	6,921 33.6%
1% to 9% (Increase)	2,621 3.8%	636 3.0%
10% to 24%	5,437 7.9%	1,551 7.5%
> 25%	6,150 8.9%	3,377 16.4%
Total	69,177 100%	20,580 100%

Amounts in thousands.

Table II.5
Distribution of Wage Changes
All Month-to-Month Main-Job Industry Changes
(Hourly Workers: February 1986 - April 1987)

Percent Change in Wages	Independent Mode (1985 Panel)	Dependent Mode (1986 Panel)
<-24% (Decrease)	3,986 10.4%	2,399 19.5%
-24% to -10%	3,367 8.8%	1,653 13.4%
9% to -1%	6,378 16.7%	1,507 12.2%
0% (No Change)	16,709 43.8%	2,692 21.8%
1% to 9% (Increase)	3,668 9.6%	1,400 11.3%
10% to 24%	2,172 5.7%	1,472 12.0%
> 25%	1,919 5.0%	1,206 9.7%
Total	38,191 100%	12,332 100%

Amounts in thousands.

Table II.6
Distribution of Employer Changes
All Month-to-Month Main-Job Industry Changes
(All Workers: February 1986 - April 1987)

	Independent Mode (1985 Panel)	Dependent Mode (1986 Panel)
No Change in Employer	50,822 73.5%	3,578 17.4%
Change in Employer	18,355 26.5%	17,003 82.6%
Total	69,177 100%	20,580 100%

Amounts in thousands.

III Concentration of Occupational and Industrial Change

The definition of occupational and industrial change used in Section II simplifies the computation of concurrent changes in related measures. It does, however, over simplify empirical reality in that changes between main and secondary jobs are counted as real occupational and/or industrial changes, when in fact they may simply reflect changes in work hours for individuals with two concurrent jobs. In this section we will be concerned with the effects of mode of data collection on the occupational (and to a lesser extent, industrial) distribution of change. To the extent that there are differences in the occupational distribution of second jobs, ignoring second jobs may bias our results. A more sophisticated definition of employment transitions is, therefore, required.

The types of employment transitions we are concerned with here are most easily seen by reference to a couple of actual event histories from the sample. Two such histories are presented in Figure III.1. The first example is that of a janitor (Occupation 453) employed in an elementary or high school (Industry 842) from the beginning of the observation period (January 1986) through August 1986. From September through November 1986 the individual had no employment. Beginning in December 1986 and continuing at least through April 1987 the individual became a houseman (Occ 449) in a hotel or motel (Ind 762).

Throughout this section we will be concerned with changes in what we will term the 'reference occupation and industry'. By this we will mean the main (WS1) occupation and industry at the beginning of an occupational or industrial spell (or the beginning of the observation period). To be conservative we will define a spell as ending when the reference occupation or industry is completely abandoned and a new main one taken on. By this definition, the first event history in Figure III.1 indicates that the individual was a school janitor from January through **November** 1986 (albeit from September through November he was an unemployed school janitor). In December a new occupation and industry spell as a hotel houseman began and continued throughout the remainder of the observation period. The individual had a single change in occupation and industry over the period.

Figure III.1
Sample Wage-Salary Event Histories

Case 1

1986/87	Ja	Fe	Mr	Ap	My	Jn	Jy	Au	Se	Oc	No	De	Ja	Fe	Mr	Ap
WS1 EID	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1
WS1 OCC	453	453	453	453	453	453	453	453	0	0	0	449	449	449	449	449
WS1 IND	842	842	842	842	842	842	842	842	0	0	0	762	762	762	762	762
SEAM	x				x				x				x			
PP INT	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
WS2 EID	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WS2 OCC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WS2 IND	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Case 2

1986/87	Ja	Fe	Mr	Ap	My	Jn	Jy	Au	Se	Oc	No	De	Ja	Fe	Mr	Ap
WS1 EID	1	1	2	2	2	2	0	0	1	1	2	2	2	2	1	1
WS1 OCC	379	379	157	157	157	157	0	0	469	469	157	157	157	157	434	434
WS1 IND	850	850	842	842	842	842	0	0	850	850	842	842	842	842	641	641
SEAM			x				x				x				x	
PP INT	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
WS2 EID	2	2	1	1	1	1	0	0	2	2	0	0	0	0	2	2
WS2 OCC	157	157	379	379	379	379	0	0	157	157	0	0	0	0	157	157
WS2 IND	842	842	850	850	850	850	0	0	842	842	0	0	0	0	842	842

The reasons we have chosen these definitions becomes clearer when we examine the second event history in Figure III.1. This is also an actual case although it is unusually complicated. The individual began the observation period doing mostly clerical work in a university (Occ 379/Ind 850). They had a second job, however, as a (substitute?) school teacher (Occ 157/Ind 842). In March of 1986 they started doing more teaching than clerking, and they did no work at all in July and August. In September, the person went back to work at both jobs, but this time the university occupation was classified as 'personal service, n.e.c' (Occ 469). In November, the individual left the university altogether and teaching became the primary job. Finally, in March of 1987, the person took a job as a bartender (434) in an eating and drinking establishment (641), and continued as a teacher as a secondary job.

If one were to look only at the first wage and salary record and were to count nonemployment as a separate occupation and industry, there would be five occupational and industrial transitions over the sixteen month period for second event history in Figure III.1 (i.e. 1: Occ 379/157, 2: 157/0, 3: 0/469, 4: 469/157, and 5: 157/434). With our more conservative definition, there are only two changes in reference occupation (379/469, and 469/157) and only one change in reference industry (850/842). In our opinion, the more conservative estimates are a better reflection of actual job stability. This person was, after all, a school teacher at the beginning and ending of the observation period and, with the exception of two summer months, during every month of the period. They did, however, leave two occupations at their university job and eventually left the university itself.

Table III.1 presents the estimated average and total number of changes in 3-digit reference occupation observed, under the conservative definition, by mode of data collection and the 40 category occupation code used in many 1990 census tables. Additionally, the ratio of independent to dependent change and the t-ratio of the difference in the average number of changes are presented in the table. The independent data collection method results in significantly more change for every occupation category. There is considerable variation by occupation in the extent to which mode of collection effects measured change. The ratio of the independent to dependent occupational change estimates is lowest for 'cashiers' and highest for 'Health Diagnostic Occupations', where no changes were observed with the dependent mode. In general, mode of collection appears to be most important for those occupations where there are high barriers to entry and consequently low levels of occupational mobility (e.g. physicians, executives, college professors, and miners). By the same token, collection mode is less important in service sector jobs where true turnover is high (e.g. 'other protective service', which includes private security guards, and 'food services' which includes fast food cooks).

Table III.2 presents the average numbers of changes in 3-digit reference industry by 2-digit industry code. Again, independent data collection results in significantly greater change for virtually every industry group. Furthermore, method of collection appears more important for industries characterized by large economies of scale (quarrying, petroleum refining, engine and turbine manufacturing, etc.), and less important for industries with low capital requirements and other barriers to entry (apparel, shoe and furniture stores, as well as eating and drinking places).

The results for both occupation and industry change are consistent with the conclusion that independent data collection results in a great deal of spurious change. The amount of this spurious change is important relative to real change in those occupations and industry where true change is rare.

Table III.1
Average Number of 3-Digit Occupational Changes
January 1986 through April 1987
by 2-Digit Occupation and Mode of Collection

2-Digit Occupation	Occupation	Independent Mode	Dependent Mode	Ratio	t-ratio of difference
1	Executive Officials and Administrators	1.324 {890}	0.116 {69}	11.414	7.016
2	Managers and Administrators	1.398 {14078}	0.197 {1773}	7.096	27.384
3	Engineers	1.673 {2726}	0.241 {310}	6.942	11.440
4	Mathematical/Natural Scientists & Architects	1.842 {1970}	0.366 {333}	5.033	9.152
5	Health Diagnosing Occupations	0.480 {163}	0.000 {0}	0.480	3.251
6	Health Assessment and Treating Occupations	0.628 {1354}	0.068 {116}	9.235	7.166
7	Teachers - Elementary and Postsecondary	0.747 {2636}	0.143 {365}	5.224	9.296
8	Postsecondary Teachers & Librarians	1.203 {2017}	0.157 {301}	7.662	11.267
9	Professionals, n.e.c.	0.871 {2641}	0.176 {441}	4.949	9.142
10	Health Technologists and Technicians	0.971 {1141}	0.107 {124}	9.075	7.531
11	Technicians, n.e.c.	1.546 {3831}	0.211 {467}	7.327	15.073
12	Supervisors and Proprietors - Sales Occupations	1.093 {2601}	0.299 {674}	3.656	9.094
13	Sales Representatives Finance and Business Services	0.912 {2329}	0.273 {622}	3.341	7.522
14	Cashiers	0.853 {2260}	0.354 {1008}	2.410	6.559
15	Sales Workers, n.e.c.	1.125 {4726}	0.357 {1395}	3.151	11.742

Population total changes in thousands in braces.

Table III.1
(continued)

2-Digit Occupation	Occupation	Independent Mode	Dependent Mode	Ratio	t-ratio of difference
16	Computer Equipment Operators	1.840 {1302}	0.262 {163}	7.023	9.475
17	Secretaries, Stenographers, and Typists	1.055 {5504}	0.256 {1140}	4.121	13.040
18	Other Financial Records Processing Occupations	1.137 {2791}	0.165 {298}	6.891	10.635
19	Mail and Message Distributing Occupations	0.706 {722}	0.175 {154}	4.034	4.298
20	Administrative Support Occupations	1.544 {14991}	0.291 {2473}	5.306	26.138
21	Private Household Occupations	0.771 {727}	0.229 {232}	3.367	4.166
22	Police and Firefighters	0.775 {591}	0.130 {86}	5.962	4.526
23	Other Protective Service Occupations	0.701 {578}	0.277 {205}	2.531	3.243
24	Food Service Occupations	1.124 {6475}	0.415 {2215}	2.708	11.535
25	Health Service Occupations	0.811 {1534}	0.219 {450}	3.703	7.376
26	Cleaning and Building Service Occupations	0.946 {3264}	0.212 {706}	4.462	10.531
27	Personal Service Occupations	0.765 {1361}	0.244 {380}	3.135	5.304
28	Farm Operators and Managers	1.225 {209}	0.200 {77}	6.125	5.370
29	Agricultural, Forestry and Fishing Occupations	1.057 {1849}	0.269 {498}	3.929	7.515
30	Mechanics and Repairers	1.179 {5038}	0.214 {752}	5.509	13.483

Population total changes in thousands in braces.

Table III.1
(continued)

2-Digit Occupation	Occupation	Independent Mode	Dependent Mode	Ratio	t-ratio of difference
31	Construction Trades	1.023 {3787}	0.213 {728}	4.803	11.218
32	Extractive Occupations	1.815 {436}	0.113 {20}	16.062	4.275
33	Precision Production Occupations	1.424 {5491}	0.262 {925}	5.435	15.221
34	Machine Operators and Tenders, except precision	1.468 {8512}	0.246 {1150}	5.967	18.308
35	Hand Working Occupations & Production Inspectors	1.255 {3482}	0.210 {540}	5.976	12.491
36	Motor Vehicle Operators	0.982 {3658}	0.246 {674}	3.992	9.508
37	Other Transportation Occupations	1.115 {306}	0.105 {21}	10.619	3.201
38	Material Moving Equipment Operators	1.314 {1255}	0.169 {148}	7.775	7.499
41	Miscellaneous Manual Occupations	1.547 {7826}	0.261 {1192}	5.927	20.044
99	Other	0.763 {495}	0.203 {156}	3.759	4.093
100	Total	1.188 {127543}	0.244 {23366}	4.869	68.735

Population total changes in thousands in braces.

Table III.2
Average Number of 3-Digit Industrial Changes
January 1986 through April 1987
by 2-Digit Industry and Mode of Collection

2-Digit Industry	Industry	Independent Mode	Dependent Mode	Ratio	t-ratio of difference
1	Agricultural Production - Crops & Livestock	0.715 {951}	0.225 {321}	3.178	4.883
2	Horticultural & Agricultural Services	1.072 {558}	0.301 {168}	3.561	3.687
4	Metal & Coal Mining; Petroleum & Natural Gas Extraction	0.732 {630}	0.200 {119}	3.660	3.381
5	Nonmetallic Mining and Quarrying	1.394 {178}	0.162 {18}	8.605	3.741
6	Construction	0.472 {2682}	0.182 {890}	2.593	6.385
10	Meat and Dairy Products, Canned Fruits and Vegetables	0.662 {623}	0.203 {151}	3.261	3.479
11	Grain Mill, Bakery, and Sugar Products	0.990 {509}	0.121 {32}	8.182	3.702
12	Beverage Industries & Miscellaneous Food Preparations	0.959 {457}	0.236 {97}	4.064	3.482
13	Tobacco Manufacturers & Knitting Mills	0.749 {130}	0.098 {12}	7.643	2.181
14	Floor Coverings and Yarn, Thread and Fabric Mills	0.944 {568}	0.312 {182}	3.026	3.450
15	Apparel and Accessories & Fabricated Textile Products	0.618 {934}	0.112 {128}	5.518	5.151
16	Pulp, Paper, & Paperboard Mills and Boxes	1.278 {799}	0.000 {0}	1.278	7.684
17	Printing, Publishing & Allied Industries	0.707 {1224}	0.105 {150}	6.733	6.243

Population total changes in thousands in braces.

Table III.2
(continued)

2-Digit Industry	Industry	Independent Mode	Dependent Mode	Ratio	t-ratio (differen
18	Plastics and Synthetics, Drugs, Soaps & Cosmetics	0.942 {514}	0.080 {39}	11.775	5.17
19	Paints, Varnishes, Agricultural & Industrial Chemicals	0.640 {414}	0.241 {130}	2.656	2.781
20	Petroleum Refining and Products; Coal Products	1.138 {232}	0.000 {0}	1.138	3.654
21	Tires and Inner Tubes, Plastic Footwear and Belting	1.349 {924}	0.214 {158}	6.304	7.603
23	Logging, Sawmills and Millwork; Wood Buildings	0.667 {375}	0.087 {59}	7.667	4.686
24	Furniture and Fixtures, Miscellaneous Wood Products	0.468 {325}	0.217 {123}	2.157	2.030
25	Glass, Cement, and Structural Clay Products	0.601 {248}	0.014 {7}	42.929	3.862
26	Pottery & Misc. Nonmetallic Mineral and Stone Products	1.198 {229}	0.147 {32}	8.150	5.082
27	Blast Furnaces, Iron, Steel, and Aluminum Industries	0.740 {404}	0.117 {85}	6.325	4.688
28	Cutlery, Hardware & Fabricated Structural Metal Products	1.136 {1015}	0.147 {83}	7.728	6.045
29	Screw Machine Products & Metal Forgings/Stampings	1.504 {457}	0.273 {86}	5.509	4.774

Population total changes in thousands in braces.

Table III.2
(continued)

2-Digit Industry	Industry	Independent Mode	Dependent Mode	Ratio	t-ratio of difference
30	Misc. Fabricated Metal Products & Nonspecified Metal Industries	1.329 {637}	0.292 {87}	4.551	4.283
31	Engines and Turbines; Farm and Construction Machinery	1.023 {407}	0.055 {18}	18.600	4.890
32	Metal working, Accounting, and Computing Machinery	1.267 {1290}	0.205 {190}	6.180	7.762
33	Nonspecified Machinery & Machinery Except Electrical	1.184 {1033}	0.257 {254}	4.607	6.313
34	Household Appliances; TV & Communication Equipment	1.061 {2464}	0.178 {358}	5.961	10.427
35	Motor Vehicles and Aircraft Including Equipment and Parts	0.754 {1291}	0.082 {136}	9.195	7.580
36	Ships, Locomotives and Space Vehicles Including Parts	0.602 {393}	0.183 {92}	3.290	2.763
37	Scientific and Controlling Instruments & Optical Supplies	1.116 {855}	0.282 {165}	3.957	4.777
38	Photographic Equipment, Watches & Clocks	0.814 {474}	0.240 {180}	3.392	4.676
39	Toys, Sporting Goods and Misc. Manufacturing Industry	0.701 {380}	0.072 {29}	9.736	3.595
40	Railroads & Bus and Taxicab Services	0.196 {190}	0.029 {16}	6.759	2.247

Population total changes in thousands in braces.

Table III.2
(continued)

2-Digit Industry	Industry	Independent Mode	Dependent Mode	Ratio	t-ratio of difference
41	Trucking, Warehouse, and U.S. Postal Services	0.357 {948}	0.150 {369}	2.380	3.993
42	Water and Air Transportation & Pipelines	0.500 {427}	0.054 {31}	9.259	3.712
43	Services Incidental to Transportation	0.620 {173}	0.410 {98}	1.512	0.857
44	Radio and TV Broadcasts & Telephone and Telegraph Services	0.485 {692}	0.129 {141}	3.760	3.816
46	Electrical Power and Gas Supply Systems	0.533 {483}	0.000 {0}	0.533	5.391
47	Water Supply and Sanitary Services	1.171 {539}	0.050 {8}	23.420	3.818
50	Motor Vehicles, Furniture, and Construction Materials	1.486 {542}	0.402 {202}	3.697	5.476
51	Metals & Minerals; Sporting and Electrical Goods	1.107 {393}	0.256 {118}	4.324	4.093
52	Hardware, Plumbing and Heating Supplies	1.176 {381}	0.187 {45}	6.289	3.848
53	Machinery; Scrap and Waste Materials	1.362 {1655}	0.434 {415}	3.138	6.693
54	Paper Products, Drugs, Chemicals & Fabrics	1.560 {515}	0.227 {61}	6.872	4.514
55	Groceries; Farm and Petroleum Products	0.840 {929}	0.211 {178}	3.981	5.004
56	Alcoholic Beverages & Farm Supplies	1.237 {801}	0.243 {104}	5.091	5.503

Population total changes in thousands in braces.

Table III.2
(continued)

2-Digit Industry	Industry	Independent Mode	Dependent Mode	Ratio	t-ratio of difference
58	Lumber and Hardware Stores & Retail Nurseries and Garden Stores	0.722 {512}	0.323 {172}	2.235	2.788
59	Mobile Home Dealers & Department and Variety Stores	0.490 {1306}	0.241 {616}	2.033	3.800
60	Grocery and Dairy Product Stores	0.532 {1589}	0.215 {694}	2.474	5.544
61	Retail Bakeries & Motor Vehicle Dealers	0.620 {895}	0.150 {191}	4.133	5.125
62	Automobile and Home Supply Stores; Gasoline Service Stations	0.917 {853}	0.300 {218}	3.057	4.256
63	Apparel, Shoe, and Furniture Stores	0.608 {867}	0.294 {378}	2.068	3.310
64	Appliance and Drug Stores; Eating and Drinking Places	0.574 {3663}	0.289 {1709}	1.986	6.103
65	Liquor, Sporting Goods, and Book Stores	0.968 {541}	0.346 {205}	2.798	3.861
66	Jewelry and Sewing Stores, Mail Order Houses	0.594 {127}	0.119 {29}	4.992	2.386
67	Vending Machine Operators; Fuel and Ice Dealers	1.308 {524}	0.178 {77}	7.348	6.043
68	Retail Florists, Miscellaneous Retail Stores	0.895 {705}	0.196 {129}	4.566	4.817
70	Banking, Credit Agencies, Savings and Loans	0.425 {1181}	0.195 {425}	2.179	3.704

Population total changes in thousands in braces.

Table III.2
(continued)

2-Digit Industry	Industry	Independent Mode	Dependent Mode	Ratio	t-ratio of difference
71	Investment Companies, Insurance and Real Estate	0.472 {1867}	0.149 {508}	3.168	6.050
72	Advertising; Services to Dwellings and Other Buildings	0.815 {621}	0.261 {230}	3.123	4.069
73	Commercial Research; Business Consulting Services	1.013 {1430}	0.338 {535}	2.997	7.239
74	Computer Services; Detective and Protective Services	0.965 {1932}	0.315 {485}	3.063	6.577
75	Electrical Repair & Automobile Services and Repair Shops	0.719 {808}	0.205 {192}	3.507	4.372
76	Private Households & Hotels and Motels	0.561 {1457}	0.231 {589}	2.429	5.253
77	Laundry, Cleaning Services & Beauty Shops	0.306 {300}	0.136 {155}	2.250	2.819
78	Barber Shops, Funeral Services & Shoe Repair	0.450 {47}	0.192 {17}	2.344	0.679
79	Dressmaking Shops & Miscellaneous Personal Services	0.924 {275}	0.290 {98}	3.186	2.889
80	Theaters, Bowling Alleys & Misc. Recreation Services	0.793 {1048}	0.219 {261}	3.621	5.608
81	Offices of Physicians	0.686 {568}	0.091 {76}	7.538	5.407
82	Offices of Dentists, Chiropractors and Optometrists	0.455 {221}	0.175 {86}	2.600	2.081
83	Hospitals & Nursing and Personal Care Facilities	0.358 {2173}	0.101 {561}	3.545	7.324

Population total changes in thousands in braces.

Table III.2
(continued)

2-Digit Industry	Industry	Independent Mode	Dependent Mode	Ratio	t-ratio of difference
84	Legal Services & Elementary and Secondary Schools	0.285 {2392}	0.076 {526}	3.750	7.470
85	Colleges and Universities; Libraries and Vocational Schools	0.466 {1402}	0.206 {668}	2.262	4.689
86	Job Training & Child Day Care Services	0.809 {720}	0.160 {133}	5.056	4.802
87	Residential Care Facilities; Museums and Zoos	0.997 {1400}	0.313 {450}	3.185	6.626
88	Religious Organizations; Engineering and Surveying Services	0.600 {1216}	0.150 {217}	4.000	5.787
89	Auditing Services and Noncommercial Scientific Research	0.743 {492}	0.167 {117}	4.449	4.680
90	Executive & Legislative Offices	1.189 {956}	0.149 {91}	7.980	7.283
91	Justice, Public Order, and Safety	0.149 {205}	0.021 {27}	7.095	3.112
92	Public Finance and Taxation; Human Resources Administration	0.675 {813}	0.105 {82}	6.429	4.392
93	Environmental and Economic Administration & National Security	0.800 {1359}	0.113 {166}	7.080	7.140
99	Member of the Armed Forces	0.742 {473}	0.203 {156}	3.655	3.943
100	Total	0.653 {70094}	0.183 {17523}	3.568	42.190

Population total changes in thousands in braces.

IV Effects of Collection Mode in Occupational Event-History Analysis

Event history analysis is one of the more promising new tools available to academic and policy analysts interested in the dynamic processes underlying employment and program participation behavior. The monthly dating of occupation and industry of employment, along with its large size and national representativeness, potentially make the SIPP one of the most attractive data bases for such analyses. The extent to which the well documented measurement errors in the SIPP (and all other panel studies with reference periods longer than the basic unit of measurement) detract from this potential is not well understood. The natural experiment resulting from the overlap in the 1985 and 1986 SIPP Panels, which differ only in the method of collecting occupation and industry, provides a unique opportunity to gain insight into the effect of measurement errors on event history analysis.

In the preceding sections of this report, we have seen that the amount of month-to-month gross-change in occupation (industry) is roughly six (four) times as great when they are asked and coded each wave (independent collection method), than when they are only asked if the respondent reports a change in duties (dependent method). In the present section we will attempt to assess the relative empirical validity of the occupation change measures for the two methods in the context of event-history analyses. To do so, we need external knowledge. Lacking direct observations of "true" occupation change, we are forced to use other forms of knowledge to serve as bench-marks in assessing the relative validity of the measures.

Our confidence in these external bench-marks varies considerably. The best source of external knowledge available to us is a consequence of the basic sample design. Specifically, the SIPP has been designed so that independent quarter samples are interviewed monthly on a rotating basis. This means that, no matter what the calendar-month pattern of true change, in the absence of measurement error, each month of the four-month reference period should contain roughly one-quarter of the observed change in occupation. The extent to which change clusters at the seams of consecutive reference periods, therefore, provides a very sound method of assessing the relative quality of the data from the two collection methods.

The second source of external knowledge upon which we can assess validity is the extent to which observed change is associated either with other observed changes or with the levels of characteristics of sample members and/or their occupations. True change in occupation "should" be related to change in wages and "should" be more prevalent for younger individuals, whereas false change need not. The extent to which one data collection method produces stronger associations (in the "right" directions) with these other measures is, therefore, a measure of its relative empirical validity. The accuracy of the assessed relative validity depends, in this case, on the accuracy of our a priori knowledge. Nevertheless, if the differences in the results obtained under the two collection modes are large, we should be able to distinguish the 'better' of the two for the purposes of event history analysis.

IV.1 Definitions of Occupational Spells

The event history models we will examine here attempt to understand the timing of individual exits from occupations as a function of the characteristics of the individuals and their occupations, as well as of the amount of time the individual has been in the occupation. This last fact means that for an observation to be informative about the dynamic process, we must know when the individual entered the occupation. When we do not observe the beginning of the spell, the spell is termed **'left-censored'**. When the end of a spell is not observed, the spell is termed **'right-censored'**. Unlike left-censored spells, right-censored spells do provide some information about the dynamic process--although not as much information as do **'completed'** spells (those in which both the beginning and end are observed).

Figure IV.1
Occupational Employment Vectors

Case 1

Occupation	Ja	Fe	Mr	Ap	My	Jn	Jl	Au	Se	Oc	No	De	Ja	Fe	Mr	Ap
453	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
449	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1

Case 2

Occupation	Ja	Fe	Mr	Ap	My	Jn	Jl	Au	Se	Oc	No	De	Ja	Fe	Mr	Ap
379	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
157	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1
469	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
434	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

To facilitate the definition of occupational spells, it is useful first to re-organize the event histories from the raw data into occupation time vectors. For each unique occupation recorded in the observation period, we can create 16x1 vectors of dummy variables equalling one if the individual reported work in that occupation (or industry) in each of the 16 months of the observation period. Figure IV.1 presents these occupational vectors for the two event histories presented earlier.

Since we cannot tell if the first occupation (453) for case 1 of Figure IV.1 began in January of 1986 or earlier, we will consider it a left-censored spell. The second occupation (449), however, clearly began in December of 1986 and continued at least through April of 1987. This occupational spell is, therefore, right-censored but not left-censored, and since we can tell that it lasted to at least its fifth month, it does provide useful information on the underlying dynamic process.

The second case depicted in Figure IV.1 is more complex and serves to illustrate the importance of assumptions regarding the length of time an individual must be absent from an occupation for it to be considered as ending. Specifically, the second occupation (157) is as a school teacher. If we were to consider any full month absence from an occupation as ending a spell, there would be two distinct spells for this occupation--one left-censored spell ending in July 1986 and one right-censored spell beginning in September. Clearly these are traditional vacation months and we would not want to define occupational spells on the basis of such temporary absences. For this reason, we will require a minimum of three-months absence from an occupation to constitute the termination of a spell. Thus, the teaching spell for case 2 of Figure IV.1 is a single spell which is both left- and right-censored.

This '**three month rule**' also has implications for the definition of censored spells. All occupations beginning before April 1986 will be considered left-censored and all spells ending after January 1987 will be considered right-censored. Finally, with the three month rule, it is still possible for there to be multiple spells for a given occupation within the sixteen month reference period.

IV.2 Effects of Collection Method on Occupational Spells

Table IV.1 presents a variety of descriptive statistics for occupational spells by method of data collection for the 1985 and 1986 SIPP panels. The average number of occupational spells per individual observed in the 16 month observation period is some seventy percent higher for the independent data collection method (2.39) than dependent (1.47).[2] This is consistent with the results described in the preceding sections. Similarly, the amount of censoring is substantially higher with the dependent collection method than with the independent. More than two-thirds (66.8%) of the occupational spells observed in the 1986 panel (dependent collection) were left censored as opposed to less than three-fifths (56%) in the 1985 panel (independent collection). Even more dramatic is the increase in right censoring and dual censoring (i.e. both left and right censoring) brought about by the dependent data collection method. Right-censored spells are nearly fifty-percent (65.1% vs. 43.9%) more prevalent with dependent data collection than with independent, while dual censored spells are more than twice as prevalent (43.8% vs. 21.3%). Thus, one effect of data collection method on event history analysis is to reduce the number of <u>apparently</u> informative (i.e. non left-censored) spells available for analysis--from 4867 for the 1985 Panel to 3578 for the 1986.

[2]In our preliminary analysis of a five-percent random subsample, the number of spells was only forty percent higher with the independent than with the dependent collection method. The reason is that their analysis ignored secondary spells in a given occupation whereas our full-sample estimates take these spells into account.

Table IV.1
Occupational Event History Statistics
by Mode of Collection

	Independent Mode (1985 Panel)	Dependent Mode (1986 Panel)
Sample Size (Weight-Sum in Thousands)	11,042 (107,766)	10,304 (96,115)
Average Number of Occupational Spells January 1986 - April 1987	2.39	1.47
Percent of Spells Left Censored	56.0	66.7
Percent of Spells Right Censored	43.9	65.1
Percent Left <u>and</u> Right Censored	21.3	43.8
Percent of Non-Left Censored Spells Beginning at Seam	82.0	52.5
Percent of Non-Right Censored Spells Ending at Seam	87.1	64.5
Percent of Completed Uncensored Spells Beginning and Ending at Seam	67.0	23.6

Whether or not the reduced number of non-left-censored spells brought about by dependent data collection represents a loss of information is not clear since we do not know what the true pattern of occupational change is. We do know, however, that in the absence of measurement error entrances and exits from occupations should be evenly spread over the months of the reference period. Furthermore, given the SIPP design we would expect approximately one-quarter of all occupational spells to begin (end) at a seam and only about a sixteenth to both begin and end at seam months. The extent of clustering of transitions at the seam months is, therefore, an indication of the extent of respondent errors in properly placing events in time. Table IV.1 indicates that, in this respect, the dependent data collection method provides cleaner data. Roughly eighty-five percent of all non-left-censored spells observed with the independent collection mode ended at a seam month. This compares with less than sixty-four percent with the dependent collection mode. Even more revealing is the fact that more than two-thirds of the completed spells observed with the independent collection mode **both began and ended** at the seam. This compares with slightly less than a quarter of those observed with the dependent collection method.

Thus, while the number of seam-coincident occupation changes is higher than it should be for either mode of data collection, it is substantially closer to what it should be with the dependent mode.

IV.3 Effects of Collection Method on Event History Model Estimates

While the above descriptive statistics suggest that the dependent data collection method results in less but cleaner data regarding the timing of entrances and exits from occupations, it does not necessarily follow that parameter estimates for event history models will be significantly affected by collection mode. To investigate this issue, we must first develop an explicit event-history model and then compare the estimates obtained from the two collection methods. Because the appropriate treatment of left-censored spells depends crucially on whether the hazard rate is a function of time in occupation or industry, we will include time explicitly in our formulation.

Specifically, we will investigate models based on the following hazard function:

$$h(t_i) = \Lambda(\alpha + \beta' X_i + \gamma_1 t_i + \gamma_2 t_i^2 + \gamma_3 S(t_i))$$

where Λ is the logistic function ($\Lambda(z)=\exp(z)/(1+\exp(z))$), X_i is a vector of characteristics of individual i, β is a vector of effects of these characteristics on the hazard of exiting an occupation, t_i is the time the individual has been in the occupational spell, and $S(t_i)$ is a dummy variable equaling 1 if month t_i is a 'seam' month.

Allowing for right-censoring, the likelihood function for our model can be expressed as:

$$L = \prod_{t_i < t_{max}} f^i(t_i) \prod_{t_i \geq t_{max}} (1 - F^i(t_i))$$

where t_{max} is the right limit of the observation period, $f^i(t_i)$ is the probability density of individual i exiting at time t_i **given** that he/she has not exited prior to that time, and $F^{i}(t_i)$ is the corresponding cumulative density function. The first product in equation 2) represents the contribution to the likelihood function of the non-right censored spells, while the second portion represents that of the right censored spells. The cumulative density function is related to the hazard function by:

$$F(t_i) = \prod_{t=1}^{t_i-1} (1-h(t))$$

where h(t) is the hazard of exiting at time t. Similarly, the probability density function is related to the hazard function via:

$$f(t_i) = h(t_i) \prod_{t=1}^{t_i-1} (1 - h(t))$$

Substituting equations 3) and 4) into equation 2) yields the following likelihood function for the discrete time model:

$$\begin{aligned} L &= \prod_{t_i<t_{max}} h(t_i) \prod_{t=1}^{t_i-1} (1-h(t)) \prod_{t_i \geq t_{max}} \prod_{t=1}^{t_i-1} (1-h(t)) \\ &= \prod_{t_i<t_{max}} h(t_i) \prod_{i} \prod_{t=1}^{t_i^{*}-1} (1-h(t)) \end{aligned}$$

where $t^{*} = t_{max}$ for right censored cases.

When dealing with a Non-EPSEM sample such as the SIPP, one can apply sampling weights via the following weighted-likelihood function:

$$L_w = \prod_{t_i<t_{max}} h(t_i)^{w_i} \prod_{i=1}^{n} \prod_{t=1}^{t_{max}^{*}-1} (1-h(t))^{w_i}$$

where w_i is the individual's sampling weight scaled by the average weight of the sample.

Finally, equation 6) is made a function of α, β and γ by substituting equation 1) for $h(t_i)$, and consistent estimates of these parameters can be obtained by maximizing the natural logarithm of the result with respect to them.[3] While we shall maximize the logarithm of equation 6) directly using an algorithm written by the author, we should note that it is also possible to use packaged logit programs by creating t_i pseudo-observations for each of the i individuals in the sample (see e.g. Allison, 1984). Also, we should note that as the number of time periods in the observation period increases, the probability of an individual exiting in any one period decreases and that, in the limit, our model reduces to Cox's proportional hazards model.

Table IV.2 presents the estimates obtained by maximizing equation 6) with respect to the parameters of equation 1) for the combined 1985-1986 SIPP panels as well as for each panel separately. The sample consists of the first non-left censored occupational spell observed for each individual.[4]

The question as to whether mode of data collection has a significant effect on event history analyses is a very straight forward one with a clear formal test procedure. Under the null hypothesis of no structural difference, the likelihood-ratio statistic $-2(\ln(L_c)-\ln(L_i) - \ln(L_d))$, where subscripts c, i and d represent combined, independent and dependent data collection respectively, is distributed χ^2, with degree of freedom equal to the number of parameters in the model. In our case, this statistic is 1,178 with 12 degrees of freedom and the null hypothesis is clearly and soundly rejected. Thus, method of data collection makes a big and very highly significant difference in the estimates attained for event history analysis of occupational exits in the SIPP data. Which method yields the 'better' estimates, however, is a question

---

[3]The estimates are fully efficient only under the assumption of simple random sampling. The estimated sampling errors from maximizing equation 6) do not reflect the effects of departures from simple random sampling in the SIPP design and will tend to understate the true sample variability. Since the 1985 and 1986 SIPP designs are quire similar, however, comparisons of the estimated standard errors are still indicative of the relative precision of the estimates.

[4]We discard left-censored spells because we expect the hazard rate to be a function of time in occupation and these spells are uninformative. Non-left censored spells subsequent to the first are potentially informative but require one to assume independence between spells--a very strong assumption. If the independence assumption is violated, the use of multiple spells per individual will result in biased parameter estimates.

which is not so easily addressed. It requires examination of the individual estimated effects and some judgements regarding their 'reasonableness'.

The independent variables included in our model can be divided into three groups according to how firm our a priori knowledge is about their true effects on occupational exit hazards. The first group consists of whether the month in question is a seam month, and the occupational coding inconsistency index.[5] **In the absence of measurement errors, these variables should have no effect on the exit hazards.** The second set of a independent variables consist of time (and its square) in the occupation. While time in occupation should affect the exit hazard rate, our a priori's about the precise pattern of this effect are not very strong. The final set of independent variables are the substantive measures of characteristics of the individual (age, wage, education, gender) and the occupation (specific vocational preparation[6]). These variables should affect exit hazards with higher hazards for younger low wage individuals in occupations with little specific human capital.

---

[5]This index is taken from Jabine and Tepping's "Controlling the Quality of Occupation and Industry Data" and is the proportion of variation attributable to response error.

[6]Specific vocational preparation is a measure of the amount of time required to become proficient in an occupation. It was merged via a 3-digit occupation code match with information published in **Jobs, Work and Occupations**, (Washington: National Academy of Sciences, 1981).

Table IV.2
Discrete-Time Event History Analysis
Exit from First Non-Left Censored Occupation
(SRS t-ratios in parentheses)

	Combined Sample	Independent Mode (1985 Panel)	Dependent Mode (1986 Panel)
Constant	- 3.10** (-19.05)	- 3.38** (-15.84)	-1.37** (-4.76)
Time in Occupation	0.63** (20.55)	0.63** (14.42)	0.35** (6.93)
Time-Squared	- .08** (-26.12)	- .08** (-18.57).	- .06** (-10.39)
Whether Seam Month	3.00** (83.52)	3.58** (74.43)	1.94** (33.25)
Age (@start)	- .58** (-8.76)	- .52** (-6.12)	-1.04** (-8.45)
Age-Squared	.07** (8.16)	.06** (5.46)	.12** (7.35)
Wage	- .02** (-3.25)	- .03** (-3.07)	- .01** (-6.24)
Education	.04 (0.64)	.13 (1.55)	.01 (0.73)
Whether Black	.02 (0.44)	.05 (0.78)	- .01 (-0.11)
Whether Female	- .08* (-2.27)	- .11* (-2.50)	.02 (0.28)
Specific Vocational Preparation	- .05** (5.22)	- .03* (-2.34)	- .11** (-5.98)
Occupational Inconsistency	.05* (2.20)	- .01 (-0.39)	.14** (3.60)
Log Likelihood (base Log L)	-12,541.64 (-18,160.76)	-7,682.58 (-12,744.32)	-4,270.21 (-5,160.02)
Adj. Likelihood-Ratio Index (χ^2) d.f. = 11	30.88% (11,238**)	39.63% (10,123**)	17.04% (1,781**)
Number of Cases	10,372	6,798	3,574

Amounts in thousands.

Perhaps the most important thing to note about the estimates of Table IV.2 is that overall goodness of fit of the model, as measured by the adjusted likelihood ratio index, is substantially higher for the independent mode (39.6%) than for the dependent mode (17.0%) This is due almost entirely, however, to the gigantic effect of the seam with the independent mode data. The coefficient of 3.58 for whether the month in question is a seam month implies that odds of exiting an occupation are some 35 times (=exp(3.58)) as high in seam months than in nonseam months. The corresponding effect for the dependent mode is just under 7 (exp(1.94)).

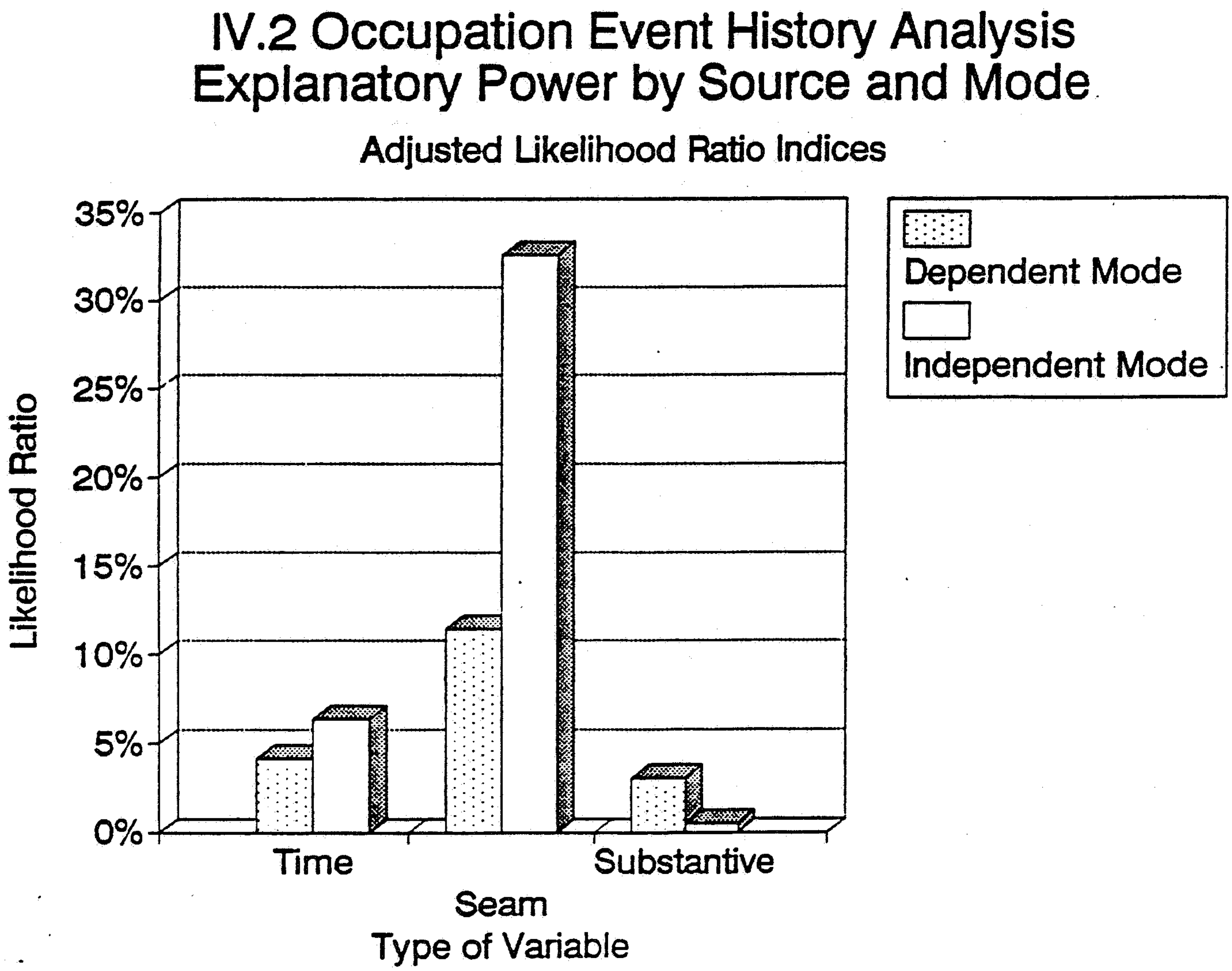
IV.4.1 Seam and Inconsistency Index

The effect of this difference in the seam variable is that it dominates the model for the independent but not the dependent mode data. Figures IV.2 through IV.4 present the marginal adjusted likelihood-ratio indices[7] for the three sets of predictors by mode of collection. Figure IV.2 shows these measures of explanatory power in absolute terms for both modes, whereas Figures IV.3 and IV.4 show them relative to the total for the independent and dependent modes, respectively. Figure IV.2 clearly shows that the seam measure (and the inconsistency index) in the independent mode model is some three times as important than in the dependent mode model, and is at least six times as important as any of the other sets of predictors.

The dominance of correlates of measurement error and the comparatively puny effects of substantive variables for occupational event history analysis using independent mode data, are even more dramatically seen in Figure IV.3. Furthermore, the improvement in the signal-to-noise ratio brought about by moving to the dependent mode of collection can be easily appreciated by comparing Figures IV.3 and IV.4. While the seam and inconsistency index still account for the majority of the overall goodness of fit with the dependent mode data, there is a clear improvement in the power of the substantive variables.

---

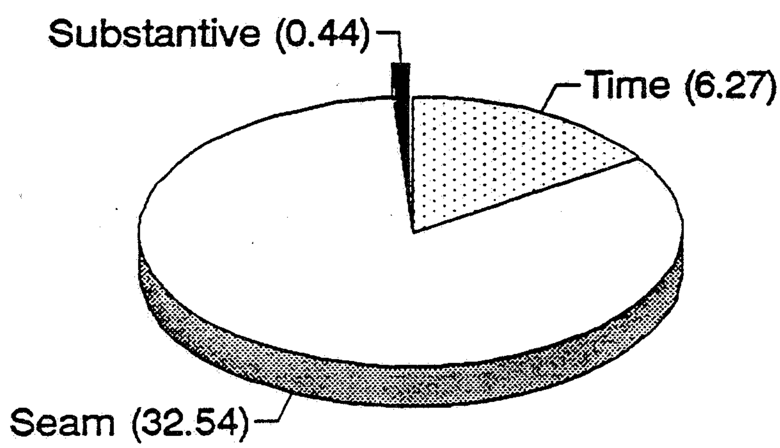
[7]The marginal adjusted likelihood-ratio index (or marginal adjusted pseudo-R^2) is obtained via $\rho^2 = (L_u - L_r + k)/L_r$ where L_r is the log-likelihood value obtained when the 'k' coefficients relating to the variables under examination are restricted to 0 and L_u is the corresponding unrestricted log-likelihood value.



IV.2 Occupation Event History Analysis
Explanatory Power by Source and Mode
Adjusted Likelihood Ratio Indices
35%
30%
25%
20%
15%
10%
5%
0%
Likelihood Ratio
Dependent Mode
Independent Mode
Time
Substantive
Seam
Type of Variable

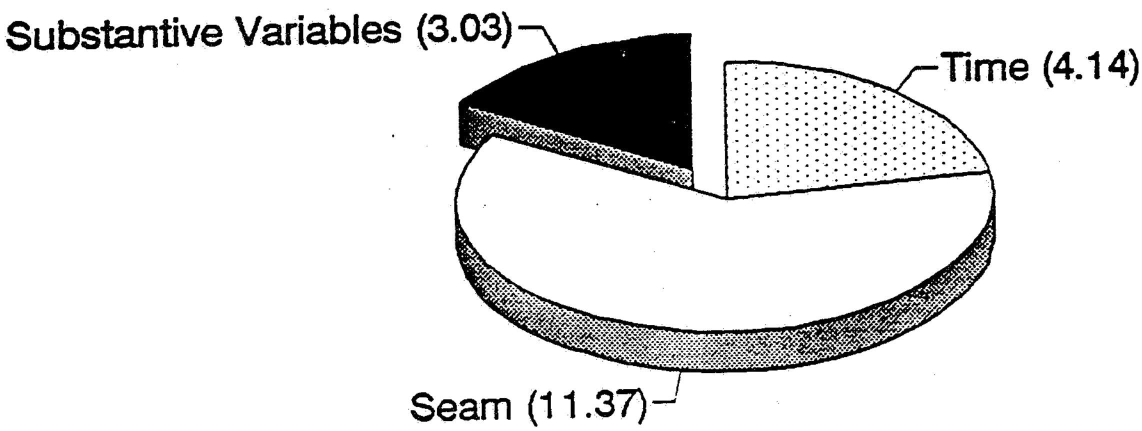
IV.3 Explained Variation
Occupation Hazard: Independent Mode

Adjusted Likelihood Ratio Indices



IV.4 Explained Variation
Occupation Hazard: Dependent Mode

Adjusted Likelihood Ratio Indices



IV.4.2 Time in Occupation

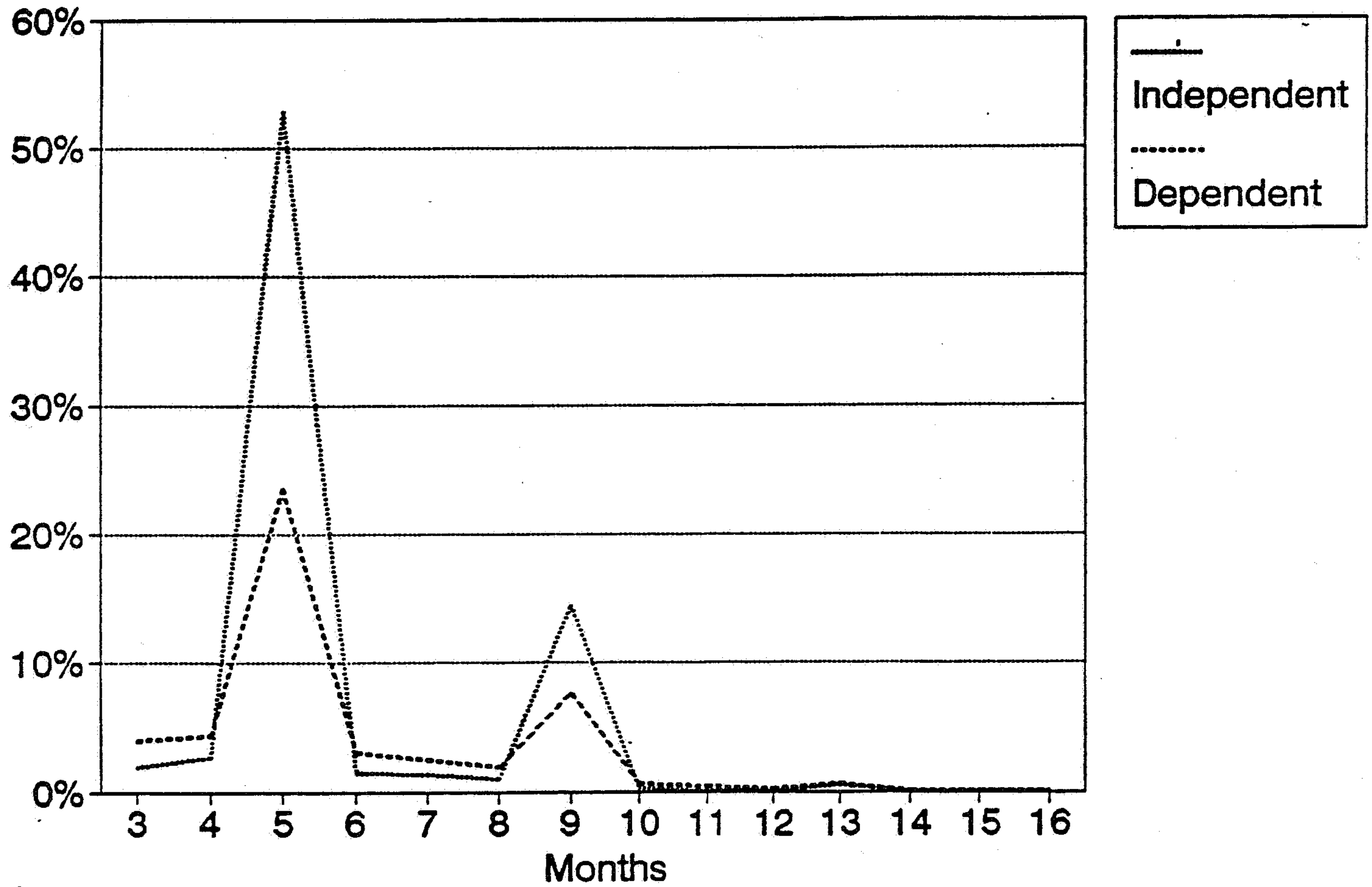
The positive and significant coefficient for time in occupation, combined with the negative (and significant) coefficient for time squared for all three samples, indicates that the hazard of exiting an occupation increases at a decreasing rate with time for the first three or four months (i.e. the function $\gamma_1 t + \gamma_2 t^2$ attains a maximum at $t = \gamma_1/(2\gamma_2) = .63/(2*0.08) = 3.9$) and declines thereafter. The combined effects of time, and whether the month of exit is a seam month, on occupational exit hazards and the occupational survival function are illustrated for members of the third rotation group in Figures IV.5 and IV.6. The effects of the seam month are apparent from the hazard functions in that they cause large spikes at the fifth and ninth months. Especially for the independent mode data these seam spikes are so dominant as to make it difficult to discern the downward trend in the base hazard--much less its acceleration with time. The corresponding estimated survival function reflects the seam effects in their step-like shape. The survival function for the independent-mode data starts out at a higher level but drops well below the dependent-mode survival function after the first seam is experienced. The shape of the survival curve for the independent mode model is, by the way, virtually identical to that obtained by Hill and Hill, 1986, for unemployment exits using a Cox proportional hazards model on the 1984 SIPP panel.

IV.4.3 Substantive Variables

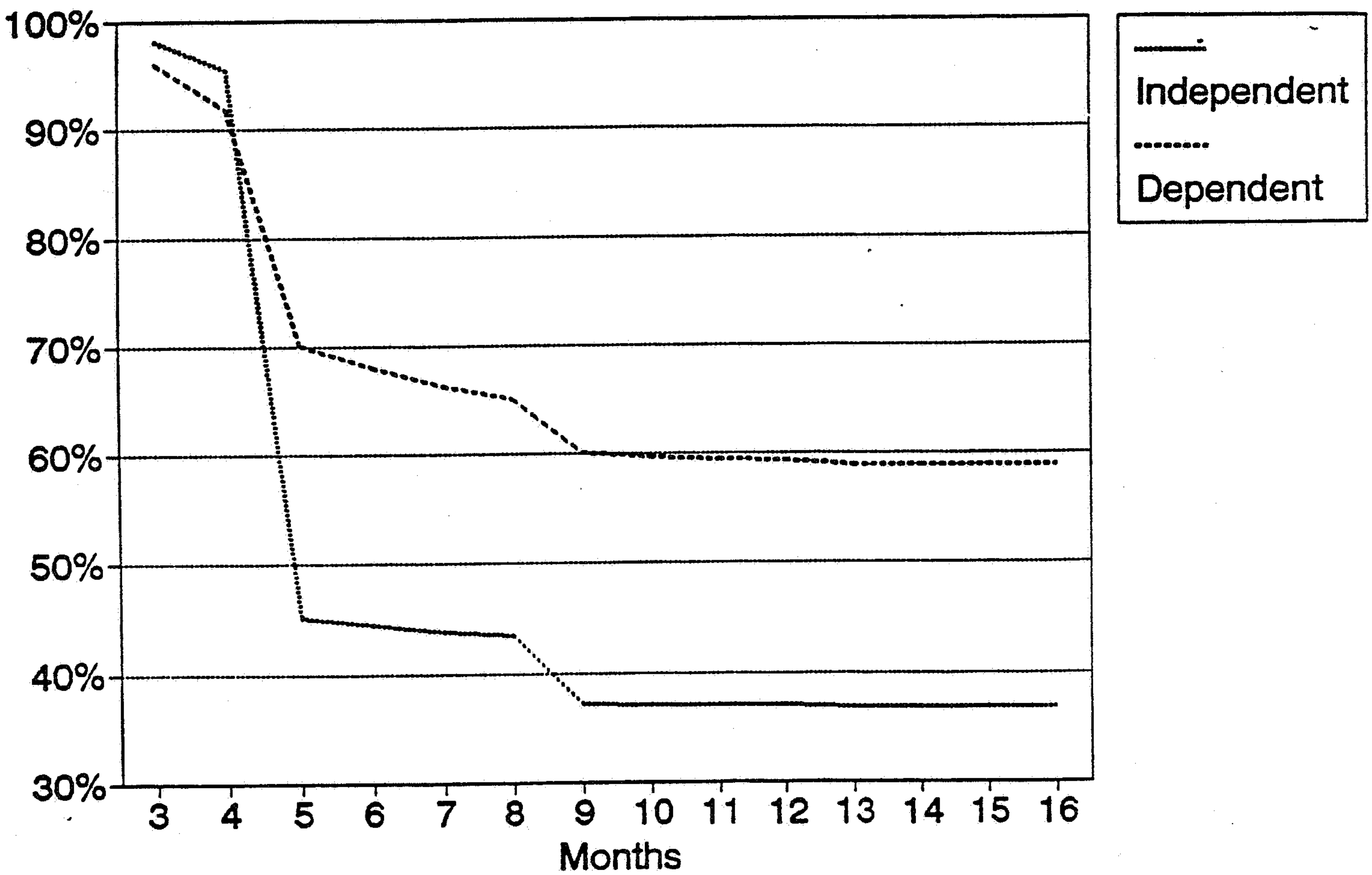
With respect to the substantive predictors, method of collection has strong impacts for estimated effects of some, but not all, measures. While the **direction** of the effects of age (at the beginning of the observation period), wage and specific vocational preparation are the same for both data collection methods, their size and significance are much stronger for the dependent mode data. For both collection modes, occupational exit hazards decrease with age at a decreasing rate until approximately age 45 at which point they begin rising again. This pattern is quite reasonable in that it reflects younger workers higher risks of unemployment and older workers higher risks of leaving the labor force. The strength of these age effects for the dependent mode data is twice that of the independent. The estimated wage and specific vocational preparation effects are also consistent with our a priori knowledge and are much more highly significant, and in the case of the specific preparation, more powerful with the dependent mode data.

The only substantive variable which is more powerful for the independent data than for the dependent is gender. The is evidence that females have lower occupational exit hazards than men in the independent data. This result seems somewhat suspect. Historically, women workers had higher exit hazards than men--a reflection of more intermittent labor force participation. By 1986, work patterns had certainly changed, with women being more likely to work even when family demands peaked, but whether these changes are enough to explain a reversal in exit hazards is doubtful.

IV.5 Occupation Exit Hazard Function by Method of Collection



IV.6 Occupation Survival Function by Method of Collection



IV.5 Effects of Occupational Characteristics on Survival

In addition to the specific vocational training measure, the Dictionary of Occupational Titles provides a number of other characteristics of occupations which may affect the hazard of leaving the occupation. These consist of three measures of job complexity ("Data Complexity"; "People Complexity" and "Things Complexity"), a measure of general educational development (GED), two measures of physical requirements (Strength and Physical Demands) and a measure of adverse environmental exposure (Environment). The most complex occupations with respect to data are those requiring the use of calculus and statistics and the least complex require only the comparison or copying of numbers.[8] The most complex "people" occupations involve either mentoring or negotiating while the simplest jobs require taking instructions. Setting up precision machine tools is the most complex "things" activity, while feeding stock to a machine is the least.

As with specific vocational preparation, we would expect job complexity to lead to reduced hazards of exiting. Again, the reason is that the level of complexity is positively associated with the extent of occupation specific human capital and both employers and employees have productivity incentives to maintain and utilize it. Similar reasoning applies to general educational development (GED) necessary to perform adequately in an occupation.

Physical demands and adverse environmental working conditions, on the other hand, should be associated with increased hazards of exiting an occupation since they are associated with manual labor and low levels of job-specific human capital.

[8]As published, the complexity scales range from low values for the <u>most</u> complex to high values for the <u>least</u>. To avoid the obvious confusion of this reverse scaling, we reverse it in our analysis so that high numbers represent high degrees of complexity.

Table IV.3
Estimated Effects of Occupational Characteristics
on Occupational Exit Hazards
by Mode of Data Collection
(SRS t-ratios in Parentheses)

Occupational Characteristic	Independent Mode (1985 Panel)	Dependent Mode (1986 Panel)
Data Complexity	- .064** (-5.35)	- .218** (-11.81)
People Complexity	- .057** (-4.54)	- .171** (-8.95)
Things Complexity	.007 (0.71)	.006 (0.41)
General Educational Development	- .048** (4.25)	- .168** (-10.13)
Specific Vocational Preparation	- .085** (-4.07)	- .327** (-11.32)
Strength	.120** (4.89)	.332** (9.45)
Physical Demands	.089** (4.11)	.245** (7.55)
Environment	.122** (4.12)	.298** (7.16)
Prestige	- .046** (-5.86)	- .135** (-11.48)
Sample Size	6798	3574

*Significant at the 95% level. **Significant at the 99% level.

Finally, occupations can be rated according to their socio-economic prestige or Duncan score. Individuals, in what the designers of the scheme consider the most esteemed occupations (e.g. Sociologists), are assigned a score of 100 while those in least esteemed occupations receive a score of 1. For a variety of reasons we would expect turnover (hence exit hazards) to be lower in high prestige occupation than in low.

Not surprisingly, there is considerable colinearity in these various scales--so much so as to preclude their inclusion as a group in our event-history models. We can, however, enter them singly in a specification which includes only the time-varying covariates (time, time-squared and whether a seam month) and compare their predictive power under the independent and dependent data collection modes. Table IV.3 presents the resulting estimates.

For all occupational characteristics other than the "Things" complexity measure (which is insignificant with either collection mode), the dependent data collection mode produces significantly stronger associations. The point estimates of the effects of occupational characteristics on exit hazards for the dependent mode are several times as large as for the independent mode. This is consistent with there being more error in occupational classification with independent collection, and this results in greater measurement error in the occupational characteristics data, with the result being attenuation of the classical sort.

Conclusions

In this section, we have attempted to assess the relative quality of occupational data obtained from the independent and dependent modes by examining the association of measured occupational change with exogenous variables. This 'empirical validity' was found to be significantly higher for the dependent (1986 SIPP Panel) mode data than for the independent (1985 SIPP Panel) mode data. Things which, in the absence of measurement error, should not effect occupational change (e.g. whether a seam month) had smaller effects with the dependent data, while things which should be associated with change had larger effects. These differences were very highly significant. Our analysis suggests that, at least within the context of event history models, the analytic potential of SIPP occupation data was increased substantially by move to dependent data collection methods.

V. Conclusions

In this report, we have examined the effects of data collection mode on the observed occupational and industrial change in the SIPP. We have done so using three distinct definitions of change and a combination of univariate, bivariate and multivariate methods. For all definitions of change and all methods of analysis, we find evidence consistent with the following conclusion:

1. the amount of gross change in occupations and industries is several times greater when the questions are asked and coded each wave (independent-mode collection) as when they are only asked if the respondent reports a change in duties (dependent-mode collection); but

2. most of the change "missed" by the dependent-mode collection methodology is noise; and

3. enough of the real change is captured by the dependent-mode collection methods as to substantially improve the signal-to-noise ratio as indicated by the higher empirical validity of the dependent-mode data.

The first two parts of this conclusion are consistent with earlier research on collection mode for occupation in the Current Population Survey (i.e. U.S. Department of Commerce, 1975). The final part of our conclusion, however, is not consistent with the finding of the 1975 Job Mobility Study that most (of what expert occupation coders considered) 'real' change in occupation is missed by dependent-mode collection methods. Both our conclusions here and those of the Job Mobility Study are judgmental. Our conclusion is based on the observation that occupational and industrial changes from the dependent-mode data relate more closely to factors they <u>should</u> relate to (e.g. changes in hours, wages, and employers and levels of age, education, and occupational characteristics) and less strongly to factors that, in the absence of measurement error, they <u>should not</u> (e.g. whether the change occurred at a 'seam' month). The judgements in the Job Mobility Study are based on whether the recorded descriptions of duties differed sufficiently in content or order of presentation to consider the two reports to be of different occupations. While the coding perspective on real change may be most appropriate for some descriptive purposes, ours is more appropriate to judging the utility of occupational and industrial change data for structural analysis.

VI REFERENCES

Allison, P.D., Event History Analysis: Regression for Longitudinal Event Data, Beverly Hills: Sage, 1984.

Cain, P.S., "The Fourth Edition Dictionary of Occupational Titles: Structure and Content", Jobs, Work and Occupation, Washington: National Academy of Sciences, 1981.

Hill, M., and Hill, D., "Labor Force Transitions: A Comparison of Unemployment Estimates from Two Longitudinal Surveys", Proceedings of the Section on Survey Research Methods, American Statistical Association, 220-225.

Institute for Social Research. (1989). A Panel Study of Income Dynamics: Procedures and Tape Codes 1986 Interviewing Year. (Wave XIX, A Supplement, Volume 1). Ann Arbor: The University of Michigan.

Jabine, T.B., and Tepping, B.J., "Controlling the Quality of Occupation and Industry Data", The Bulletin of the International Statistical Institute, 45(3), 1973.

Mathiowetz, N.A., 1985. "The Problem of Omissions and Telescoping Errors: New Evidence from a Study of Unemployment", Proceedings of the Section on Survey Research Methods, American Statistical Association, 450-454.

Neter, J., and Waksberg, J., "Response Error in Collection of Experimental Data vs Household Interviews: An Experimental Study," U.S. Bureau of the Census Technical Papers, No. 11, U.S. Government Printing Office, Washington, DC, 1965.

U.S. Department of Commerce, Social and Economic Statistics Administration, 1975, Comparison of Month-to Month Changes in Industry and Occupation Codes with Respondent's Report of Change: CPS Job Mobility Study, (Response Research Staff Report No. 75-5).