

Continuation of NLS Discussion Paper 92-4  
Part 3 of 3

This version of the paper was split for web delivery.

## APPENDIX A: Construction of a Data Set from the YNLS Describing Earnings and Employment Experiences

In this appendix the procedures used to construct data on earnings and employment are described. These data are used both to describe the earnings and employment experiences of youth, and in the subsequent analysis of unemployment insurance.

The YNLS data are briefly discussed. There follows a very detailed discussion of data derivation. Then the samples used in Section 3 are defined and discussed. It is established that while in any calendar year earnings data are usually available for at least one job, earnings data are often not available for all jobs. This is not so critical for Section 3, where incomplete earnings data may lead to only specific calendar years being dropped from the analysis. It is very critical for the unemployment insurance analysis which requires that earnings data be complete in all calendar years. Since earnings data are most often missing on less important "intermittent" jobs, these missing data are imputed by methods described at the end of this appendix.

The discussion is quite detailed. To understand the body of this appendix, it is sufficient to read only Sections A.1 and A.5, which are self-contained.

### A.1 *General Considerations*

The data used come from the first seven rounds of the National Longitudinal Survey of Youth (YNLS). This survey commenced in 1979, when the respondents were between 14 and 22 years old. Calendar year data can be constructed for the seven years from 1978 to 1984. In this study the first year, in which there is relatively more missing data, is not analyzed.

To avoid confusion about whether a quoted year is the calendar year under investigation, or the interview year from which data is obtained, interviews are referred to by the survey round number rather than the year. Survey Round 1 was conducted in 1979, Survey Round 2 in 1980 and so on.

The primary data source is the YNLS Data Tape. The version used is the public-use multi-file format tape for rounds 1 to 7 combined. In addition, the YNLS Work History Tape is used. This is also publicly available, but this study uses a pre-release version which differs from the public-release Work History in the following minor respect. To conserve space the

public-release version omitted data for the sixth through tenth job held since the previous interview. The pre-release version has data on all ten jobs. (In the public-release version, the A and DUALJOB arrays still use information on up to ten jobs, but to get data on jobs 6 through 10 will require use of the ADDJOBS files, which are available for most but not all years. At each interview between 25 and 47 of the 12,686 cases have more than five jobs.) For geographical information, the Geocode tapes are used.

The complete sample of 12,686 people includes non-random samples of the poor and military personnel. However, this study uses only the 6,111 people in the cross-section sample, designed to represent the noninstitutional civilian segment of youth aged 14 to 21 as of January 1, 1979. Otherwise there are few restrictions on the sample, and every effort is made to keep sample sizes as large as possible. In particular, only minimal attempts are made to clean or omit suspect data. One exception is that start and stop dates for jobs and gaps within jobs are checked for validity.

#### A.2 *Work History Data*

The analysis is based on data on each job held since the last interview. This is obtained in the Employer Supplements and the "On Current Labor Force Status" section of the main questionnaire. This data is stored in a convenient form in the YNLS Work History Tape, which has weekly activity arrays that give codes for every employer in that week, or if the respondent is not employed the major activity. Weekly data on labor supply and earnings can be obtained.

These data are aggregated to form calendar year data. The calendar year data will typically use data from the two interviews that together span that calendar year. For example, calendar year 1981 computed earnings will generally use data from both the 1980 and 1981 interviews (but will use other interviews if one or both of these interviews are missed).

A major task is to determine which respondents should be omitted due to missing data. The term "missing" data encompasses the following situations:

- (1) Data cannot be obtained due to noninterview.
- (2) Received income from military service.
- (3) Data cannot be obtained due to age 15 years or less (every job is obtained only for

respondents 16 years or older).

- (4) Job dates given are inconsistent; e.g. stop before start.
- (5) Data cannot be obtained due to falling into a class that is not asked the relevant questions; e.g. wage rates are not obtained for all jobs. (These are coded as -4 in the raw data.)
- (6) Missing in the sense used in the YNLS: the respondent was asked the question but refused to answer (-1) or did not know (-2), or should have been asked the question but was not (-3).

Most data are "missing" for the first five reasons.

#### A.2.1 *Data on Each Job in Each Week*

For all respondents 16 years and over, detailed information is obtained for every "regular" civilian job held since the previous interview. This information includes the dates the job started and stopped; within this, dates for periods during which the respondent did not work for the employer; hours per week usually worked at the job; and the usual wage rate. Unfortunately, the wage rate is not obtained for every job.

In many instances a respondent will be working for the same employer at different interviews. A code exists to link jobs with the same employer across interviews. To avoid confusion the following terminology is used. Data on a job is obtained directly from the job data at each interview. Data on an employer is obtained by linking different job entries from different interviews for the same employer. Most of the analysis in this study is done at the job level, i.e. the fact that different jobs from different interviews may be with the same employer is ignored.

The key variables are based on the following definitions, drawn from the YNLS questionnaire:

Job:

"Some jobs are odd jobs - that is, work done from time to time, like *occasional* lawnmowing or babysitting. Others are *regular* jobs - that is, jobs done on a more or less regular basis. (Not counting the job you had last week), Since (DATE OF LAST INTERVIEW), have any jobs you've had for pay been done on a more or less regular basis? Please give me the names

of each of your employers for all *regular* jobs you've had for pay since (DATE OF LAST INTERVIEW) (not counting the job you had last week)."

Additional questions are asked to ensure recording of all jobs for pay with government sponsored programs such as college work-study, high school cooperative work-study, Neighborhood Youth Corps In-School, summer employment, and employer tax credit.

The job last week is the job picked up in the "On Current Labor Force Status" section of the main questionnaire, where questions virtually the same as those in the monthly CPS are asked. In particular, the respondent is asked to report any work at all last week, not counting work around the house, and to give details for the employer with whom the respondent worked the most hours last week. This job may be either occasional or regular. (About 5 percent of the CPS jobs are occasional rather than regular in Surveys Round 3 to 6 for the full sample of 12,686).

Note, for the items below, the respondent is asked "For *all* of the *rest* of the questions we have about (EMPLOYER), please think only of the time you worked for (EMPLOYER) since (DATE OF LAST INTERVIEW).

#### Gaps within Jobs:

"For one reason or another, people often do not work for a week, a month, or even longer. For example, strikes, layoffs, and extended illnesses can cause people to miss work for a week or longer. Between (DATE STARTED JOB / LAST INTERVIEW) and (DATE STOPPED JOB / NOW), were there any periods of a full week or more during which you did *not* work for this employer, not counting *paid* vacations or *paid* sick leave?"

Up to 4 such gaps are reported.

#### Hours per week:

"How many hours per week (do/did) you *usually* work at this job?"

## Wage:

"Altogether, including tips, overtime, and bonuses, how much (do/did) you *usually* earn at this job? Please give me the amount you earn(ed) *before deductions* like taxes and Social Security (are/were) taken out. Was that per hour, per day, per week, or what?"

The wage is reported as hourly, daily, weekly, bi-weekly, monthly and annual. The wage rate question is only asked if one or more of the following conditions are satisfied: (1) The job is the current job recorded in the CPS section. (2) The job is part of a government-sponsored program. (3) The job has been held for more than 9 weeks and is for 20 or more hours per week and the respondent is 16 years and older. Thus wages for intermittent jobs are missing.

The necessary jobs-related data from the Employer Supplements are stored on both the original raw YNLS data set, and the YNLS Work History Tape. The latter is used here as the data are stored as an easily accessed PL/I data structure. In addition, the Work History tape has useful constructed variables such as the A and DUALJOB arrays described below. The Work History tape is accompanied by documentation that includes a listing and description of the PL/I program that created the work history tape.

The work history program uses the raw data on start and stop dates for jobs, and start and stop dates for gaps within jobs, to construct weekly activity arrays, called the A and DUALJOB arrays, which detail every job that the respondent had that week. (For those who had no civilian job, additional data from the "Military" and "Gaps when R was not Working or in the Military" sections are used, and the arrays indicate whether the person was in the active armed services, or unemployed or out of the labor force).

The dates are originally entered to the day. Employment in any day of a week is treated as employment for the whole week. For example, a job that begins on a Wednesday is treated as beginning on the previous Sunday, and a job that ends on a Wednesday is treated as ending on the following Saturday. If job start and stop dates are randomly distributed across the week, the length of employment at each job will on average be overstated by a week. However, the bias is nowhere near as great as this for the following reasons. A disproportionate number of jobs begin on a Monday and end on a Friday. For jobs from

different survey rounds associated with the same employer, the problem arises only for the start date of the job in the first survey that the employer is recorded and the stop date of the job in the last survey that the employer is recorded. The start and stop dates for gaps within jobs are similarly treated, which imparts a potential bias in the opposite direction.

A detailed description of the program that constructs jobs data from the YNLS Work History Tape follows. Access to documentation for this data set is assumed.

#### Hourly Wage and Weekly Wage for Each Job:

If the job is not for pay, the Work History program sets HOURLYWAGE to -4. This needs to be recoded to zero. If CLASSWORKER equals 4 and (PAYRATE equals 0 or TIMERATE equals -4) then HOURLYWAGE equals 0. For Rounds 1 to 7 there are a total of 105 such jobs.

The weekly wage is constructed in the obvious fashion. If HOURLYWAGE  $\geq 0$  and HOURSWEEK  $> 0$  then WEEKLYWAGE equals HOURLYWAGE times HOURSWEEK. This algorithm can compute WEEKLYWAGE only if HOURSWEEK is reported. In some cases it is possible to construct WEEKLYWAGE even if data on HOURSWEEK are missing. This is the case when wages are reported as weekly (then WEEKLYWAGE = PAYRATE), bi-weekly (then WEEKLYWAGE = PAYRATE / 2), monthly (then WEEKLYWAGE = PAYRATE / 4.3) and yearly (WEEKLYWAGE = PAYRATE / 52). These calculations are done only in those cases where HOURSWEEK is missing. For rounds 1 to 7 there are a total of 105 such jobs. WEEKLYWAGE is truncated to the nearest cent.

In the Work History program, HOURLYWAGE is truncated to the nearest cent. This makes no difference if wages are reported at an hourly rate, the case for half the reported wages. But for wages reported as daily, etc. the hourly wage and weekly wage will be slightly understated.

#### Missing Data Because Not Interviewed:

If the respondent misses one or more rounds of the survey, but is interviewed at a later round, data are not missing, since all the necessary questions are asked for the period since

the last interview. If the respondent is not interviewed at a later round, then data are set to missing for weeks subsequent to the week of the last interview.

#### Missing Data because Active Military Service:

The Work History Tape includes start and stop dates for each period of active service in the military. Active service is service in the branches coded 1 to 4 in the "Military" section of the questionnaire; viz. army, navy, air force, marine corps. It does not include any of the Reserves or National Guards. The dates are used rather than a code of 7 in the A array, as when a person in the military also holds a civilian job the A array records the job rather than military service. For weeks in which the respondent is in the active services, data are set to missing.

#### Missing Data because Bad Dates:

The constructed data are based on the A and DUALJOB arrays. These only include jobs for which valid start and stop dates are available. If the dates are invalid, there is no record of the job in the A or DUALJOB array, and no indication that the job is missing. Similarly, if the dates for gaps within jobs are invalid, there is no indication of the gap, and no indication that the gap is missing. (Though for the first gap there is a record, the A array being set to 3). For weeks in which dates for jobs or gaps within jobs are invalid, annual computed earnings are set to missing. Also, more stringent tests of date validity are used.

A respondent is treated as having a job if  $START > -4$  or  $STOP > -4$ , and having a gap within jobs if  $WEEKSNOTWORKED \neq 0$  and  $WEEKSNOTWORKED \neq -4$ . Dates for jobs are invalid for the following reasons:

- (1)  $START > STOP + 1$  or  $START < 0$  or  $STOP < 0$ .
- (2)  $START < LASTINT - 1$  or  $STOP > INT + 1$ .
- (3)  $PERIODSTART > PERIODSTOP$  or  $PERIODSTART < 0$  or  $PERIODSTOP < 0$ .
- (4)  $PERIODSTART < START - 1$  or  $PERIODSTOP > STOP + 1$ .

In the first round, some of the dates for gaps within jobs are associated with the wrong job. This is an error in the YNLS Data Tape that will be detected by the above tests (in many cases), and the additional reasons for rejection:



(5)  $PERIODSTART \geq 0$  or  $PERIODSTOP \geq 0$  when  $WEEKSNOTWORKED$  equals 0 or -4

(6)  $PERIODSTART \geq 0$  or  $PERIODSTOP \geq 0$  when  $START = -4$  or  $STOP = -4$ .

The Work History program does only checks (1) and (3). The presence of the "+1" or "-1" terms in the above tests may at first seem strange. It is necessary because of the way dates are treated for jobs held at the interview date. For example, if the respondent ended a job on 1/10/79 and was interviewed on 1/12/79, then  $STOP = CEIL(375/7) = 54$  and  $INT = FLOOR(377/7) = 53$ , in which case  $STOP > INT$  even though the dates are obviously valid. See "Description of the NLSY 1979-1985 Work History Program" and the program itself, for further details on its treatment of dates.

If the data fail the checks above, job dates are treated as being invalid from  $LASTINT$  to  $INT$ , or in the case of checks (4) and (5) from  $START$  to  $STOP$ . Data are set to missing for the weeks that these dates lie in. Typically two calendar years will be effected.

#### Missing Data because of Missing Wage:

As already noted, wage rates are not obtained for all intermittent jobs. The percentage of jobs for which the wage rate was deliberately not requested ranges from 29 percent in Round 1 down to 19 percent in Round 7. In these cases the job is one held for less than 9 weeks and for less than 20 hours per week. For an additional 2 percent of all jobs, wage rates are missing due to refusal, don't know, invalid skip, or code 7 for time unit rate of pay.

#### A.2.2 *Weekly Earnings and Work Experiences*

Weekly hours (WH) and weekly earnings (WE) from all jobs this week are obtained by summing usual hours per week and earnings per week over each job recorded in the weekly A and DUALJOB arrays. Hourly earnings this week ( $WE/WH$ ) is simply computed as WE divided by WH.

In section 3 variation in WH, WE and  $WE/WH$  over weeks in the calendar year is studied. The calendar year is standardized at 52 weeks. Weeks are allocated to the years 1978 through 1984 using the scheme described below in Section A.2.3.

### A.2.3 Annual Computed Earnings and Work Experiences

To obtain annual work experience or earnings data the basic approach is the following. From the A and DUALJOB arrays obtain the number of weeks in each calendar year at each job. Multiplying by each job's hours per week or earnings per week and summing across jobs yields annual hours (AH) or annual computed earnings (ACE).

To make ACE comparable with annual reported earnings (ARE, defined below) the calendar year length for ACE and AH is determined by the number of work days.

When a week spans two calendar years, jobs in that week are allocated partly to each year, according to the proportion of the work week (Monday to Friday) falling into each. Weeks begin on a Sunday, with week 1 commencing on 1/1/78. The calendar years are:

- 1978: Weeks 1 to 52.
- 1979: 53 to 104 and  $0.2 \times 105$ .
- 1980:  $0.8 \times 105$  and 106 to 156 and  $0.6 \times 157$ .
- 1981:  $0.4 \times 157$  and 158 to 208 and  $0.8 \times 209$ .
- 1982:  $0.2 \times 209$  and 210 to 261.
- 1983: 262 to 313.
- 1984: 314 to 365 and  $0.2 \times 366$ .

Each calendar year is processed in turn, using the A and DUALJOB entries for the weeks in that particular year to compute the number of weeks in each job held that calendar year. Only job entries in the A and DUALJOB arrays are processed. In particular, code 3 in the A array is ignored here. It is picked up as missing data at a later stage.

For example, consider the following A and DUALJOB entries:

Weeks	150-170	171-190	191-210
A	201	301	302
DUALJOB	202	302	0

Then in calendar year 1981 there are 13.4 weeks at job 201, 13.4 weeks at job 202, 20.0 weeks at job 301, and 38.8 weeks at job 302.

For each job held in the calendar year, annual earnings are computed in whole dollars as the product of the weekly wage and the number of weeks at the job, divided by 100 and

truncated to an integer value. Then sum over all jobs.

Continuing the earlier example suppose the weekly wage for job 201 is 8000 cents, for job 202 is 22490 cents, for job 301 is 11100 cents, for job 302 is 9500 cents, and for job 303 is 25704 cents. Then for calendar year 1981:

$$\begin{aligned} \text{Computed earnings} &= \text{FLOOR}(8000 \times 13.4/100) + \text{FLOOR}(22490 \times 13.4/100) \\ &\quad + \text{FLOOR}(11100 \times 20.0/100) + \text{FLOOR}(25704 \times 38.8/100) \\ &= 1072 + 3013 + 2220 + 9973 \\ &= \$16,278 \end{aligned}$$

For computed earnings to be comparable with the reported earnings data, the reported wage for each job in each calendar year should be the average wage for the job that calendar year.

Since the reported wage is the usual wage received over the period worked since the last interview, this will be the case on average for respondents interviewed on January 1 each year. There will be no bias.

For respondents interviewed at other times, the reported wage for each job will not be the average for the calendar year. But this will not induce any biases. To see this, consider the following simple example. The respondent works for only one employer, with the weekly wage path:

	Jan-Mar	Apr-Jun	Jul-Sept	Oct-Dec
1981	\$200	\$210	\$220	\$230
1982	\$240	\$250	\$260	\$270
1983	\$280	\$290	\$300	\$310

The average weekly wage for calendar year 1982 is \$255, and reported earnings will be 52 times \$255. Suppose the respondent is always interviewed on March 31. Then at the Round 4 survey on March 31, 1982 he should report a usual weekly wage of \$225 (the average of \$210, \$220, \$230, \$240), and at the Round 5 survey on March 31, 1983 a usual weekly wage of \$265 should be reported. Computed annual earnings are 13 times \$225 plus 39 times \$265, which equals 52 times \$255, as desired. Again there is no bias.

There are clearly cases where there will be individual biases, due to jobs held only in the first few weeks after or before an interview, or interviews not in the same month each year, or interviews missed entirely. But there is no *a priori* reason to believe that these will not balance out over all jobs and respondents.

Annual weeks worked (AWW) is the number of weeks in the calendar year for which a job is recorded in the weekly A array. AEMPS is the number of employers in the calendar year, obtained by summing over jobs held in the year but not double-counting jobs with the same employer. (Recall that work for an employer may appear as two jobs this year – one from the survey this year and one from the survey next year).

Annual weeks with a multiple job (AWMJ) is the number of weeks in the calendar year for which a job is recorded in both the A and DUALJOB arrays. The work history data are cleaned up to avoid spurious double counting in the survey week – in some cases the work history data records both the entry from this survey and the entry from the next survey.

### A.3 *Annual Reported Earnings*

In the "On Assets and Income" section of the questionnaire, all respondents are directly asked the amount received from various income sources for the calendar year preceding the interview date. In this paper, annual reported earnings (ARE) are the sum of wage and salary earnings and own farm or business earnings.

Respondents are generally interviewed in the early part of the year. About half the interviews take place in January or February, and over 90 percent of interviews are completed by the end of April. The latest interview month is August. So the recall period for the reported earnings questions is not too long.

The key variables are based on the following definitions:

#### Wage and Salary Earnings:

"During 19xx, how much did you receive from wages, salary, commissions, or tips from all jobs, before deductions for taxes or anything else?" (Not counting any money you received from your military service).

#### Own Farm or Business Earnings:

"During 19xx, did you receive any money in income . . . from your own farm? from your own nonfarm business, partnership or professional practice?"

Construction of this variable is straightforward. It is the sum of the two components. The only complication is for those respondents in Rounds 1 to 4 who satisfied all of the following: under 18 years, never married, never had a child, never enrolled in college and lived at home. These respondents were asked a shorter set of income questions. Separate questions were asked for whether or not the respondent received income from (A) working on own business or farm and (B) interest on savings or any other income received periodically or regularly, not counting allowances from parents. However, these respondents were then asked the amount received from A and B combined. For these people data are treated as missing if the answer is yes to both A and B. At most 30 respondents in each of Rounds 2 to 4 are missing earnings data for this reason.

The main reasons for missing data on reported earnings are that the person was not interviewed, or that earnings from service in the military were reported. (There is a separate question asked: "Did you receive any income from service in the military?" If the response is yes or missing, reported earnings are treated as missing that year. This separate question is not asked in the shorter set of questions mentioned in the previous paragraph, but respondents asked the shorter set are unlikely to be in the military, and if they are, they may still be picked up as such in the computed earnings section).

The amount received from each source of income is truncated at \$75,001 for rounds 1 to 6, and at \$100,001 for Round 7. This occurred for 1 respondent in Round 4, and 4 respondents in each of Rounds 5, 6 and 7. In these instances, data are treated as missing. Note that it is still possible for reported income to exceed \$75,001 (or \$100,001) if each component is not truncated. Data are also missing if the person was (erroneously) not asked or did not reply to questions on either or both of the components of reported income. Most of this was due to the respondent not knowing the amount received from wages and salary.

#### *A.4 School Attendance and Education Level*

Data on school attendance and educational level can be constructed for each calendar year. However, the school data are much more complete from the Round 3 survey on. In

particular, before 1980 it can be determined that a respondent did not attend school at any time in the year, but for those who did attend at some time the length of school attendance cannot be determined without further assumptions.

The school data come directly from questions in the "Regular Schooling" section about attendance since the preceding interview at regular school: elementary school, middle school, high school, college or graduate school. Some questions about other types of schools and training programs are asked elsewhere but are not used here.

Monthly school attendance data are often, but not always, available. When monthly data are available a weekly attendance array is created by assuming that attendance in any month means attendance for every week (beginning Sunday) that falls in the month. When monthly data are unavailable, use the date last enrolled in school. For weeks after this date and before the current interview the respondent is not in school, while for weeks prior to this date and after the preceding interview, school attendance is uncertain. Any attendance or uncertain attendance during a calendar year leads to exclusion of youth for that calendar year from the analysis in Section 3.

The level of education is based on the question: "What is the highest grade or year of regular school that you have completed and gotten credit for?" This is for all youth in survey Round 1, and in subsequent rounds for all youth that at any time since the last interview attended or were enrolled in regular school. In addition, separate questions on attainment of high school diploma and attainment of college degree are used to increase education to 12 years (high school diploma), 16 years (bachelors degree) and 17 years (masters degree) where appropriate. A question on generalized equivalency degree (GED) is *not* used. Such youths will be assigned less than 12 years of education, unless they obtain further schooling.

#### A.5 *Variables and Samples Used in Section 3*

The following variables, for each individual in each calendar year or each week, are the basic data for Section 3.

Annual Earnings data:

ACE\_ = Annual Computed Earnings from all jobs for which wage data is available

ACE = Annual Computed Earnings from all jobs. Constructed only if wage data is available for all jobs

ARE = Annual Reported Earnings (March CPS type question)

ARE\_ = Annual Reported Earnings constructed only if ACE > 0.

Annual Work Experience data:

DAEMP = Dummy for whether employed or not at any time during year

AWW = Annual Weeks Worked

AEMPS = Number of Employers over the year

AH = Annual Hours at all jobs

AH\_ = Annual Hours at all jobs for which wage data is not missing

ADMJ = Dummy for whether or not simultaneously held more than one job in any week this year

AWMJ = Annual weeks held multiple job, given held a multiple job.

Weekly Earnings and Weekly Work Experience data:

WE = Weekly earnings from all jobs this week

WH = Weekly Hours from all jobs this week.

From these basic variables, we additionally construct:

ARE/AWW = Weekly Reported Earnings

ACE/AWW = Weekly Computed Earnings

ARE/AH = Hourly Reported Earnings

ACE\_/AH\_ = Hourly Computed Earnings

AWMJ/AWW = Percentage of annual weeks with multiple job, given held a multiple job

WE/WH = Weekly Hourly Wage from all jobs this week.

Finally for the weekly data WE, WH and WE/WH we construct Average, Max, Min, Relative Range (RR) and Absolute Range (AR) for the weekly data WE, WH and WE/WH. These refer to Average, Max, Min, Relative Range and Absolute Range for a given individual

over the year across weeks with non-zero non-missing data. As an example consider weekly earnings:

$$\text{AVE (WE)} = (\text{Sum of WE over weeks with non-zero non-missing WE and WH})/\text{AWWP}$$

$$\text{Max (WE)} = \text{Maximum of WE over weeks with non-zero non-missing WE and WH}$$

$$\text{Min (WE)} = \text{Minimum of WE over weeks with non-zero non-missing WE and WH}$$

$$\text{RR (WE)} = \text{Ln (Max (WE)/Min (WE))}$$

$$\text{AR (WE)} = \text{Max (WE)} - \text{Min (WE)}.$$

All earnings and wage data are inflated to 1984 constant dollars by the All Items CPI for urban consumers (Economic Report to President 1989 Table B-58).

Construction of these variables is described in Sections A.2-A.4. A crucial distinction is between ACE and ACE\_. Annual computed earnings from all jobs, ACE, can be constructed only if earnings data are available for all jobs held in the calendar year. Annual computed earnings from all jobs for which wage data is available, ACE\_, can be constructed if earnings data are available for at least one job. The sample sizes for ACE\_ will be considerably greater than those for ACE.

ACE is used for analysis of annual earnings. Since the sample for ARE, annual reported earnings, is much larger, we additionally define ARE\_, annual reported earnings given ACE can be constructed, to permit comparable samples for analyzing computed and reported earnings.

ACE\_ is used for analysis of hourly earnings to compute hourly computed earnings, given earnings for at least one job, we divide ACE\_ by AH\_, annual hours at all jobs for which wage data are available.

Regarding sample compositions used in the analysis, the empirical work in Section 3 is limited to youth age 18 years or more, not in the military and not in school at any time during the year, and with education of grade 8 or more. To ensure cell sizes of at least 30, look at:

AGE 18-19	ED 8-11, 12	Years 79-84
20-22	8-11, 12, 13-15	79-84



23-24	8-11, 12, 13-15	81-84
25-27	8-11, 12, 13-15, 16+	83-84

where

AGE = age in years at March 12 of the calendar year.

ED = highest grade of completed schooling at the end of the calendar year.

Beyond that, whenever data is available it is used. Since the number of missing observations varies by data items, this leads to many samples.

Sample sizes are given in Tables A.1-M, for men, and A.1-W, for women. For each cell, the upper three entries are for samples A, B, and C; the second three entries are for samples D, E and F; the third three entries are for samples G, H and I; and the lowest three entries are for samples J, K and L. Samples A to M are defined in Table A.2. A listing of the samples used for each of the variables in the tables in Section 3 is given in Table A.3.

The analysis of Section 3 is for all youth not in the military or in school (sample A), leading to a sample initially larger for women. This sample is used to compute the employment rate - DAEMP. Youth are dropped from the analysis only to the extent that relevant data are missing.

The primary data throughout this study are the work history data on each job held since the preceding interview. The analysis of this data in section 3 is restricted to those who worked at some stage during the year (sample B). This and subsequent samples are smaller for women than for men, except for youth with 12 years of education.

To analyze weeks worked during the year and related variables - AWW, AEMPS and ADMJ - we need to exclude youth for whom the dates of employment for any job are missing or invalid (sample C). About 2 percent of youth are excluded for this reason, mostly in 1979 due to missing data in the first survey, and more often male. Sample C is the basis for all subsequent samples but sample H.

TABLE A.1-M  
SAMPLE SIZES a/.b/

GRADE	AGE	1979			1980			1981			1982			1983			1984			1979-84 ALL		
8-11	18-19	114	108	96	131	117	114	151	135	133	165	144	140	116	96	93	46	43	43	723	643	601
		94	95	86	114	113	108	133	130	124	142	141	136	93	91	92	43	42	41	621	614	590
		84	98	86	107	110	102	124	125	118	137	130	126	91	89	84	41	39	39	591	551	533
		95	55	51	101	67	60	117	75	65	125	94	82	80	63	56	38	24	20	556	370	324
20-22		102	99	89	159	150	145	177	159	156	199	175	172	241	211	209	262	241	239	1140	1031	1010
		89	82	83	145	145	139	156	155	154	172	171	164	209	205	205	239	237	235	1110	945	980
		85	86	77	139	128	123	154	143	137	164	162	154	206	193	188	234	229	226	963	841	810
		72	58	44	125	101	85	138	123	108	154	127	119	183	166	151	222	170	164	903	746	674
23-24								59	55	53	89	77	75	105	95	95	127	119	119	380	346	341
								52	52	50	74	74	73	95	94	91	119	119	117	340	329	311
								50	49	46	73	68	65	90	89	86	117	117	116	320	323	313
								47	40	36	66	56	49	86	70	66	114	90	89	313	256	240
25-27														54	48	47	96	88	86	150	136	133
														47	47	45	86	86	85	133	133	130
														45	43	41	85	85	81	120	128	122
														41	32	29	82	65	63	123	97	92
12	18-19	174	173	155	177	167	164	174	173	170	164	160	158	154	142	141	87	82	82	930	897	870
		155	152	151	164	160	160	170	166	167	158	157	157	141	141	140	82	82	78	870	859	853
		150	158	141	157	155	149	165	160	159	154	153	150	138	132	131	78	77	76	842	835	805
		153	92	85	146	107	99	157	121	114	150	116	109	131	96	84	75	52	52	812	584	545
20-22		249	246	227	342	335	326	369	359	349	390	377	371	387	378	373	360	350	346	2097	2045	1992
		227	211	219	326	323	321	344	344	345	371	376	370	372	370	367	346	346	342	1991	1962	1964
		221	228	208	319	314	301	338	338	328	366	362	357	360	352	347	336	338	333	1940	1932	1875
		209	143	136	303	249	233	327	269	255	356	280	268	348	281	264	330	238	230	1873	1460	1386
23-24								130	128	124	213	206	203	230	218	216	259	253	253	832	805	796
								124	123	123	203	201	202	216	214	211	253	251	250	796	785	781
								121	122	119	199	194	191	206	207	203	248	245	245	774	758	753
								120	100	99	192	156	148	203	164	158	243	174	174	757	596	579
25-27														123	118	114	223	225	225	336	343	341
														116	116	116	225	224	224	341	340	340
														114	113	111	217	216	216	331	329	327
														113	89	85	215	182	176	320	271	261
13-15	20-22	45	45	40	71	71	70	69	67	67	63	62	62	78	76	75	79	79	79	405	400	393
		40	39	40	70	69	67	66	65	64	62	62	62	75	75	72	74	74	74	391	384	384
		40	41	35	68	69	68	65	63	63	62	58	58	70	68	67	75	74	74	383	372	365
		38	28	26	67	53	52	61	49	46	57	47	45	68	50	47	74	54	53	365	281	264
23-24								43	43	43	73	71	71	78	77	77	75	75	74	269	266	264
								43	43	43	71	70	71	77	75	75	74	74	72	265	262	261
								43	43	43	69	70	70	74	75	75	68	68	68	254	256	254
								43	37	37	68	50	49	73	51	50	60	51	47	252	189	183
25-27														48	47	47	93	93	93	141	140	140
														47	47	47	93	93	92	140	140	139
														47	46	45	91	91	91	138	137	136
														45	41	40	91	70	70	136	111	110
16-20	23-24							33	31	30	72	68	68	68	67	67	76	74	72	249	241	237
								30	29	29	68	67	68	67	67	67	72	71	71	237	230	235
								29	27	26	66	68	67	66	67	67	70	74	72	231	236	232
								26	22	22	66	47	46	67	48	48	72	61	61	231	178	177
25-27														59	53	52	97	95	94	151	148	146
														52	52	52	94	94	93	146	146	145
														50	53	52	92	91	90	142	142	142
														53	42	42	91	72	70	144	114	112

a/ Samples A to L are defined in appendix A.

b/ For each education-age-year grouping: upper three entries are respectively samples A, B, and C; second three entries are respectively samples D, E, and F; third three entries are respectively samples G, H and I; and lower three entries are respectively samples J, K and L.

TABLE A.1-H  
SAMPLE SIZES a./b/

GRADE	AGE	1979			1980			1981			1982			1983			1984			1979-84 ALL		
8-11	18-19	126	82	77	140	92	91	132	91	91	127	80	79	100	60	61	51	27	27	656	424	423
		76	77	64	91	91	78	91	90	88	80	80	80	61	61	59	17	11	14	418	418	381
		65	75	69	78	82	79	87	88	85	80	74	71	59	53	52	16	18	17	385	392	373
		71	38	33	79	42	38	81	58	55	71	61	53	52	42	37	16	11	10	270	253	229
20-22		101	62	56	155	97	96	178	104	106	197	108	107	212	142	141	198	134	133	1041	631	640
		56	54	49	96	95	92	105	103	100	107	106	102	141	140	135	133	133	129	639	631	607
		50	55	47	92	86	84	99	95	93	101	98	97	135	131	129	128	120	119	601	581	569
		50	37	33	82	66	58	88	82	73	96	78	71	128	99	91	116	87	77	540	449	423
23-24								53	30	30	95	53	51	99	47	46	123	73	72	370	203	201
								29	28	28	52	52	51	46	46	45	73	73	72	200	199	196
								28	28	26	50	48	47	45	42	42	72	69	69	195	187	184
								24	17	15	47	39	36	42	31	28	69	62	59	182	149	138
25-27														51	23	23	91	52	52	142	77	77
														23	23	23	52	52	49	75	75	72
														23	22	21	49	49	45	72	71	66
														21	19	18	44	41	37	55	60	55
12	18-19	212	193	177	243	220	228	251	232	226	217	193	192	196	173	172	82	72	72	1001	1084	1061
		178	174	172	218	218	217	227	227	227	192	192	189	172	172	169	72	72	70	1059	1055	1039
		172	181	166	210	209	205	226	219	214	184	184	183	166	161	157	72	71	71	1030	1023	998
		175	113	107	205	148	138	218	176	169	182	134	127	159	122	114	71	54	53	1010	747	708
20-22		287	240	226	414	346	339	477	356	349	464	390	380	475	401	401	478	409	404	2545	2143	2110
		226	216	212	343	340	330	349	349	345	382	382	377	401	401	396	406	405	395	2107	2093	2058
		217	225	207	329	322	315	340	334	326	375	373	360	391	386	382	392	399	393	2046	2039	1983
		209	157	143	313	253	237	330	277	260	364	311	293	380	319	307	393	301	294	1989	1618	1532
23-24								160	133	133	268	216	214	272	213	212	294	237	235	994	800	797
								133	132	131	215	212	209	212	212	211	235	235	229	795	791	780
								109	129	129	205	200	198	208	204	200	226	230	225	765	763	753
								108	103	101	196	164	152	198	157	147	226	187	182	748	611	582
25-27														160	125	125	282	218	218	442	343	343
														125	125	123	218	218	214	343	343	337
														122	120	120	211	208	206	333	328	326
														119	87	83	206	171	164	325	256	247
13-15	20-22	77	73	65	130	120	118	135	123	121	126	115	113	113	103	103	112	101	99	691	635	623
		65	62	64	116	118	116	120	120	118	112	110	109	103	103	103	100	100	99	618	613	609
		63	67	60	112	114	111	116	111	110	107	110	109	101	102	102	97	101	100	597	605	592
		64	48	45	111	77	73	110	91	85	107	85	82	102	70	69	100	64	64	554	435	410
23-24								69	63	63	106	98	98	123	110	110	109	101	101	407	373	371
								62	62	61	98	97	96	109	109	107	99	99	99	368	367	363
								60	60	60	93	93	93	105	104	103	95	97	95	353	354	351
								59	47	46	92	83	81	102	82	76	96	76	74	349	288	277
25-27														66	56	56	126	114	113	192	170	170
														56	54	56	114	112	112	160	166	168
														56	53	52	110	109	109	166	162	161
														50	42	41	107	83	79	157	125	120
16-20	23-24							42	41	41	65	64	64	84	82	82	104	102	102	295	289	289
								41	41	41	64	64	63	82	82	82	102	102	102	289	289	288
								41	37	37	60	63	62	80	81	80	101	102	102	282	283	281
								37	27	24	62	50	49	80	53	51	102	81	81	281	211	205
25-27														53	50	50	91	84	83	144	134	134
														50	50	50	83	83	82	133	131	132
														48	48	48	80	82	80	128	130	128
														46	36	35	80	59	55	128	91	93

a/ Samples A to L are defined in appendix A.

b/ For each education-age-year grouping: upper three entries are respectively samples A, B, and C; second three entries are respectively samples D, E, and F; third three entries are respectively samples G, H, and I; and lower three entries are respectively samples J, K, and L.

Table A.2

Definitions of Samples

- A: Persons not in military and not in school this year AGE  $\geq$  18 and ED  $\geq$  8.
- B: Sample A less those who did not work at any time in the year. It is assumed that those who had jobs since the last survey, but for whom even the dates of employment are bad or missing, did work during the calendar year.
- C: Sample B less those who did work but had missing or invalid dates for one or more jobs. This is the reference sample for all the analysis, except that of Annual Reported Earnings which does not require job dates.
- D: Sample C less those for whom weekly hours (hours on all jobs held in each week) are not available for even one week of the year.
- E: Sample C less those for whom hours are missing for one or more jobs held during the year.
- F: Sample C less those for whom the hourly wage rate cannot be constructed for even one job held during the year.
- G: Sample C less those for whom hourly wage each week and weekly earnings are not available for even one week of the year.
- H: Sample A less those for whom annual reported earnings are zero or missing.
- I: Sample H less those for whom annual weeks worked are missing or zero. (The intersection of samples H and C).
- J: Sample H less those for whom annual hours worked are missing or zero. (The intersection of samples H and E).
- K: Sample C less those for whom the wage rate is missing on one or more jobs held during the year.
- L: Sample K less those for whom annual reported earnings are zero or missing. (The intersection of samples H and K).

Table A.3

## Samples used for Tables 3.1-3.6

Table Variable	Sample	Name
3.1 ARE	H	Annual Reported Earnings
ARE_	L	Annual Reported Earnings constructed only if ACE $\geq$ 0
ACE	K	Annual Computed Earnings
3.2 Log(ARE_)	L	Log Annual Reported Earnings if ACE $\geq$ 0
Log(ACE)	L	Annual Computed Earnings
3.3 ARE/AWW	I	Weekly Reported Earnings
ACE/AWW	K	Weekly Computed Earnings
AVE(WE)	G	Average Weekly Earnings
AR(WE)	G	Absolute Range of Weekly Earnings
RR(WE)	G	Relative Range of Weekly Earnings
3.4 ARE/AH	J	Hourly Reported Earnings
ACE_/AH_	F	Hourly Computed Earnings
AVE(WE/WH)	G	Average Hourly Earnings per week
RR(WE/WH)	G	Relative Range of Hourly Earnings per week
AR(WE/WH)	G	Absolute Range of Hourly Earnings per week
3.5 DAEMP	A	Employed during year
AWW	C	Weeks worked in 52 week year
AEMPS	C	Number of Employers over the year
ADMJ	C	Dummy for simultaneous job holder
AWMJ	M	Number of Weeks with multiple jobs
AWMJ/AWW	M	Fraction of weeks worked with multiple jobs
3.6 AH	E	Annual Hours
AH_	E	Annual Hours with non-missing pay
AVE(WH)	D	Average Weekly Hours
RR(WH)	D	Relative Range of Weekly Hours
AR(WH)	D	Absolute Range of Weekly Hours

To investigate variation in weekly hours within the year --AVE(WH), RR(WH) and AR(WH) - weekly hours are required for at least one week (and for all jobs in that week) in the year (sample D). Less than 0.25 percent of observations are lost for this reason.

For total hours worked over the year, - AH and AH\_ - we exclude from sample C youth for whom hours worked at any job are missing (sample E). About 1 percent of sample C is lost due to missing hours data, primarily in 1979. AH\_ is the sum of hours at jobs for which earnings are known, while AH is the sum of hours at all jobs, regardless of whether earnings are known.

Hourly computed earnings - ACE\_/AH\_ - requires data for both ACE\_ and AH\_ (sample F). Thus earnings are needed for at least one job during the year and hours are needed for the jobs used in constructing ACE. Almost 3 percent of sample C is lost.

To investigate variation in weekly earnings within the year - AVE(WE), RR(WE) and AR(WE) - and hourly earnings across weeks within the year - AVE(WE/WH), RR(WE/WH) and AR(WE/WH) - we require both weekly earnings and hourly earnings for at least one week in the year (sample F). Almost 3 percent of sample C is lost. (Note that since earnings may be reported as hourly, weekly, monthly, ... there are some cases when weekly earnings are missing but hourly are not, and vice-versa, but for simplicity we have required that both be known). For youth who hold more than one job in any week of the year, AVE(WE/WH) will differ from ACE\_/AH\_.

In addition to the earnings data from the work history, data on calendar year reported earnings - ARE - are separately available (sample H). Reporting error aside, sample H should be roughly sample B less youth with missing reported earnings. Sample H is about 6 percent smaller than sample B, and is of size comparable to samples F and G which essentially require earnings and hours data for at least one job held during the year.

To compute weekly reported earnings - ARE/AWW - requires data on both ARE and AWW (sample I). This is the intersection of samples C and H. The requirement that job dates be known leads to a loss of 2 percent of sample H (similarly about 2 percent of sample B is lost for this reason).

To compute hourly reported earnings - ARE/AH - requires data on both ARE and AH

(sample J). This is the intersection of samples H and E. The requirement that annual hours be known leads to a loss of about 1 percent of sample I.

To compute annual computed earnings at all jobs - ACE - requires earnings data for all jobs held in the year (sample K). This is much more stringent than data required for ACE\_ which requires earnings for just one job in the year. Tables A.1 indicate a significant decrease in sample size.

Because so many youth have incomplete earnings data, the samples for ACE and ARE are not necessarily comparable. A better comparison is obtained by restricting analysis to annual reported earnings given complete earnings data, ARE\_ (sample L).

Finally, in investigating multiple job holdings, AWMJ and AWMJ/AWW, attention is restricted to multiple job holders (sample M). Sample sizes are not reported in Tables A.1 since most cells are very small. Sample sizes across all years can be obtained by multiplying sample sizes for sample C by ADMJ reported in Table 3.5.

The basic samples are sample A, the universe of respondents for Section 3; sample C, respondents who were employed during the year and for whom the dates of employment are not missing or invalid (e.g. start date after stop date); sample H, those who reported earnings during the year and for whom this data is not missing; and sample K, those for whom wage data is available on all jobs held during the year. Sample K (and L) is considerably smaller than the others, because the wage rate is not asked for jobs of less than 20 hours and/or less than 9 weeks, unless the job is the main job at the time of the survey or a government-sponsored job.

#### A.6 *Imputation of Missing Earnings*

Summing over all years and age-education groups in Tables A.1, sample C has 7,285 observations for men and 7,237 observations for women, while sample K has only 5,261 observations for men and 5,298 observations for women. In any one calendar year, therefore, earnings on some jobs are missing for over a quarter of youth who work during the year. For the analysis of unemployment insurance, which requires a complete time series of earnings from 1979 or the last date of school attendance, well over half the potential sample will be lost due to missing earnings.

Since most of the jobs with missing wages are less important intermittent jobs, of less than 9 weeks duration and/or less than 20 hours per week, it seems reasonable to try to impute some of these missing wages.

Recalling that hours data are available even if the wage is missing, an obvious procedure is to assign the difference between annual reported earnings and annual computed earnings for those jobs where wage data is available to hours worked at jobs with missing wages, i.e. the imputed hourly wage is  $(ARE - ACE_)/(AH - AH_)$ . A weakness of this approach is that measurement errors in ARE and ACE\_ are greatly magnified if wages are missing for only a small fraction of hours worked, which is often the case. For example, measurement error leading to ACE\_ greater than ARE leads to a negative imputed wage.

Other sources of information are instead used. The imputed hourly wage is sequentially calculated as:

- (1) hourly wage for a job with the same employer reported in the preceding or subsequent interview (appropriately deflated or inflated)
- (2) average hourly earnings at other jobs with known earnings this survey round, provided these jobs account for more than 50% of total hours at jobs this interview
- (3) average hourly earnings at other jobs with known wages in the preceding or subsequent interview (appropriately deflated or inflated), provided these jobs account for more than 80% of total hours at jobs that interview.

Source (2) is used only if (1) is unavailable, and (3) is used only if (1) and (2) are unavailable. The inflation factors for (1) and (3) are the average wage growth for the sample, 23.0% (from surveys 1 to 2), 15.3% (2 to 3), 13.4% (3 to 4), 9.5% (4 to 5), 9.5% (5 to 6) and 11.6% (6 to 7). If data from both the preceding and subsequent surveys are available to construct (1) or (3), their average is used.

The imputed hourly wage is multiplied by actual hours per week to impute earnings per week at the job. Then weekly earnings are constructed by summing over actual or imputed earnings per week at all jobs held in the week.

This use of imputed earnings greatly increases the sample size for the analysis of unemployment insurance. Nonetheless, other criteria such as interview in all years lead to a



sample considerably smaller than the potential 6,111. It should also be clear that the sample for Section 4 onwards, which uses a work history over many years, differs from the samples for Section 3, which are for separate analyses of each calendar year.

## Appendix B

This appendix provides a description of the procedures used in the construction of the data set analyzed in Sections 4 through 10. An assessment of the reliability and accuracy of our imputed measures of UI entitlements is also presented here.

### B.1 *Sample Selection*

To obtain reasonably reliable measures of weekly earnings a stringent sample selection procedure was used to obtain a subsample of youths from the nationally-representative component of the YNLS. A youth had to satisfy 6 conditions to be included in the subsample. First, to minimize the biases that can arise from mistakes in recalling events further in the past a youth must have been interviewed in each of the first 7 years of the survey. Second, he or she must have worked at least once after January 1979. Third, to assure reliable measures of time employed and nonemployed an individual was required to report valid beginning and ending dates for time periods spent working, between jobs and in the military. Fourth, he or she must have left school and not returned prior to the January 1985 interview date. Fifth, a youth must have a reasonably accurate and complete time series of either reported or imputed weekly earnings beginning in January 1978 or the last date of school attendance. Finally, the respondent must have started a nonemployment spell after March 1979 or the last date of school attendance.

Table B.1 provides a summary of the number of youths affected by each of the successive screens. The resulting subsample of 3,028 individuals from the 6,111 youths in the nationally-representative component of the YNLS is used for all of the empirical analyses in Sections 4 through 10.

### B.2 *Imputation of UI Entitlements*

The detailed work histories available in the YNLS provide a unique opportunity to construct accurate measures of the amount of UI benefits available to nonemployed youths. Every State determines an individual's eligibility to receive UI and the amount of benefits he or she is entitled to collect on the basis of the reason for leaving the latest employer and a detailed earnings history over a recent 52 week period, termed the "base period." While there

TABLE B.1  
Effect of Sample Selection Criteria on Sample Size

Screen	Number Eliminated	Number Remaining
Nationally-Representative Sample of YNLS		6111
Missed Interview	711	5400
Never Worked	185	5215
Invalid Dates	982	4233
In School	745	3488
Incomplete Weekly Earnings History	321	3167
Always Employed	139	3028

is variation across States in the specific earnings history collected and the definition of the base period, the complete time series of weekly earnings available from the YNLS is sufficient to calculate all of the earnings measures used by States to determine UI entitlements.

To establish whether a youth was disqualified from receiving UI because of the reason for separation from his or her last employer, we utilized the self-reported reason for the initiation of a nonemployment spell. Respondents were provided a wide array of possible causes for starting a period of nonemployment. We have condensed this range of responses into 8 reasons for the beginning of a period not working. Briefly, these 8 causes are: (1) on layoff; (2) discharged; (3) quit for other than family or health reasons; (4) quit to join the armed forces; (5) quit for family or health reasons; (6) quit to attend school; (7) on strike; (8) unknown or other reasons. Approximately 25 percent of the nonemployment spells began because of a layoff, another 10 percent resulted from discharges and 20 percent started after a quit for other than family or health reasons. Quits for family or health reasons and other reasons account for almost all of the remaining spells.

All States have disqualification provisions for voluntarily leaving work without good cause, discharge for misconduct<sup>38</sup> and direct involvement in a labor dispute. While the majority of State statutes do not directly specify what constitutes "good cause," most States operationally define this provision to include only causes involving the fault of the employer or other employment related reasons. As noted in Section 4, to determine the sensitivity of our results to different interpretations of the voluntary separation provision we have adopted two methods to determine eligibility based upon reason of separation. The narrow definition of eligibility disqualifies individuals unless they were on layoff or were discharged. The broad concept of eligibility disqualifies a youth if he or she reported the nonemployment spell started because of causes 5, 6, 7 or 8 listed above. This broad interpretation of the voluntary separation provision is used in all of the analyses in Sections 5 through 10.

In addition to satisfying the separation from work condition, an out-of-work individual must also demonstrate a "permanent attachment" to the labor force by attaining either a minimum level of earnings or a minimum number of weeks of work in covered employment,

---

<sup>38</sup> Our eligibility imputation procedure does not account for the misconduct provision because we are unable to distinguish between discharges for misconduct and other discharges.

or possibly both during the base period. As a result of the Federal Unemployment Tax Act virtually all employment has been covered by the UI system since 1977. The major exclusions to coverage are self-employed individuals, agricultural workers, paid participants in a government financed training program, employees of immediate family members, and certain officers of private corporations. Thus, in constructing the time series of weekly earnings in covered employment we excluded self-employment income, income from a family farm or business, income from government sponsored training programs and earnings from jobs where a youth reported his or her occupation as a farmer or farm laborer.

In conjunction with the laws of each State, the information on the reason for the initiation of a nonemployment spell and the constructed weekly time series of earnings in covered employment enabled us to impute both UI eligibility and the amount of benefits available to a youth at the beginning of each spell of nonemployment. The accuracy of these imputed measures of eligibility and UI entitlements is the subject of the last section of this appendix.

### B.3 Construction of Work History Variables

All States use some combination of average weekly earnings throughout the base period (*AWE*), highest earnings during any calendar quarter of the base period (*HQE*) and total earnings over the base period (*BPE*) to determine an individual's eligibility to collect UI as well as the amount of benefits available during the subsequent benefit year.<sup>39</sup> The specific rules and regulations determining eligibility and entitlements vary from State to State and involve complex interactions between the various earnings measures above. In addition, upper and lower thresholds in both the weekly benefit amount (*WBA*) and the number of weeks of eligibility (*WE*) introduce further nonlinearities into the relationship between entitlements and an individual's work history.

To account for the interactions and nonlinearities relating program rules and the three earnings measures, we have constructed a set of dummy variables that indicate which of a series of brackets contain the combination of *AWE*, *HQE* and *BPE* associated with a youth at the beginning of a nonemployment spell. As illustrated in Table B.2 each earnings

---

<sup>39</sup> As previously noted, programs that utilize information on weeks worked (*WW*) are combining information on *AWE* and *BPE* since  $WW = BPE/AWE$ .

TABLE B.2

## Bracket Definitions for AWE, HOE and BPE

Bracket	Earnings Measure		
	AWE	HOE	BPE
1	\$0.00-\$99.99	\$0.00-\$999.99	\$0.00-\$1499.99
2	\$100.00-\$149.99	\$1000.00-\$1999.99	\$1500.00-\$3999.99
3	\$150.00-\$199.99	\$2000.00-\$3499.99	\$4000.00-\$7999.99
4	\$200.00-\$299.99	\$3500.00-\$5499.99	\$8000.00-\$14999.99
5	\$300.00 +	\$5500.00 +	\$15000.00 +

measure was divided into 5 brackets. The endpoints of each bracket correspond to the lower and upper thresholds determining UI eligibility and entitlements for the various earning measures. Clearly, it was not possible to account for all of the complexities involved without resulting in an unacceptably small number of spells associated with individuals in any one bracket. For example, the *BPE* brackets were selected to capture the variation both across States and over time in the minimum amount of total earnings necessary to become eligible for UI, as well as the minimum level of *BPE* needed to qualify for the maximum amount of benefits available. The minimum level of *BPE* necessary to qualify for UI benefits varied from \$150 in Hawaii in 1979 to more than \$3000 in 1985. The first 2 *BPE* brackets account for this lower threshold. Similarly, the upper 3 brackets in *BPE* were chosen to allow for the nonlinearities introduced by the maximum amount of benefits payable (i.e., maximum *WBA* times the maximum *WE*) under various State programs. The minimum level of *BPE* necessary to qualify for the maximum potential benefits varied from about \$4000 in Illinois in 1979 to \$21,500 in Colorado in 1985.

Brackets in *HQE* and *AWE* were chosen in a manner similar to the procedure used to select the *BPE* brackets. States which base entitlements on *HQE* have upper and lower thresholds in *HQE* equivalent to the *BPE* limits discussed above. Variation in the minimum *WBA* and maximum *WBA* thresholds influenced the choice of brackets in *AWE*. Finally, the additional eligibility requirements of *BPE* greater than 1.5 times *HQE* or a weeks of work requirement also effected the selection of the lower brackets in both *HQE* and *AWE*.

The empirical specifications in Sections 7 through 9 incorporate a set of work history controls based on these bracket definitions. Let a certain combination of *BPE*, *HQE* and *AWE* describe a worker type and define a dummy variable  $WT_i$  equal to 1 if an individual is a worker of type  $i$  or 0 otherwise. The empirical results in this report are based on the definition of 22 worker types for the men and 15 worker types for the women as defined in Tables B.3-M and B.3-W respectively.

#### B.4 Accuracy of Imputed Measures of UI Eligibility and Entitlements

The self-reported measures of UI receipt available in the YNLS provide an opportunity to assess the accuracy of our imputed measures of UI eligibility and entitlements. While

TABLE B.3-M  
Definition of Work History Controls for Men

Worker Type	Earnings Brackets			Worker Type	Earnings Brackets		
	BPE	HQE	AWE		BPE	HQE	AWE
WT <sub>m1</sub>	1	1-2	1	WT <sub>m12</sub>	3	2	2
WT <sub>m2</sub>	1	1-2	2	WT <sub>m13</sub>	3	2	3
WT <sub>m3</sub>	1	1-2	3	WT <sub>m14</sub>	3	3	3
WT <sub>m4</sub>	1	1-2	4-5	WT <sub>m15</sub>	3	2-3	4-5
WT <sub>m5</sub>	2	1	1-2	WT <sub>m16</sub>	3	4-5	4-5
WT <sub>m6</sub>	2	2	1	WT <sub>m17</sub>	4	3	3
WT <sub>m7</sub>	2	2	2	WT <sub>m18</sub>	4	3	4-5
WT <sub>m8</sub>	2	1-2	3-4	WT <sub>m19</sub>	4	4	4
WT <sub>m9</sub>	2	3	3	WT <sub>m20</sub>	4	4-5	5
WT <sub>m10</sub>	2	3-4	4-5	WT <sub>m21</sub>	5	4	4-5
WT <sub>m11</sub>	3	2	1	WT <sub>m22</sub>	5	5	5

TABLE B.3-W  
Definition of Work History Controls for Women

Worker Type	Earnings Brackets			Worker Type	Earnings Brackets		
	BPE	HQE	AWE		BPE	HQE	AWE
WT <sub>w1</sub>	1	1-2	1	WT <sub>w9</sub>	3	2	2-4
WT <sub>w2</sub>	1	1-2	2	WT <sub>w10</sub>	3	3	3
WT <sub>w3</sub>	1	1-2	3-5	WT <sub>w11</sub>	3	3-5	4-5
WT <sub>w4</sub>	2	1	1-2	WT <sub>w12</sub>	4	3	3
WT <sub>w5</sub>	2	2	1	WT <sub>w13</sub>	4	3	4-5
WT <sub>w6</sub>	2	2	2	WT <sub>w14</sub>	4	4-5	4-5
WT <sub>w7</sub>	2	1-4	3-5	WT <sub>w15</sub>	5	4-5	4-5
WT <sub>w8</sub>	3	2	1				



the data available do not permit us to identify individual spells of UI receipt, the YNLS does provide reliable calendar year measures of the total number of weeks of UI receipt, the average *WBA* over the year and the months in which benefits were received. Thus, the following assessments, as well as the analyses in Section 4, are based upon annual measures of eligibility and entitlements constructed from our measures imputed at the beginning of a period of nonemployment.

Calculating annual values for the imputed UI variables is straight forward for individuals who experience a single spell of nonemployment that begins and ends within a single calendar year. This exercise is also relatively simple for individuals who experience multiple nonemployment spells all occurring within a calendar year. Ambiguities arise when a nonemployment spell overlaps two calendar years, especially if a person experiences more than one period of nonemployment during a given year. When this situation arose the number of weeks of eligibility were allocated to the beginning weeks of a nonemployment spell.<sup>40</sup> For example, suppose a youth was eligible for 26 weeks of UI benefits at the beginning of a 30 week nonemployment spell where 18 weeks occur in one calendar year and 12 weeks take place in the second year. In this case 18 weeks of eligibility would be assigned to the first year and the remaining 8 weeks would be allotted to the second calendar year.

Table B.4 presents a cross-tabulation of our estimated eligibility to receive benefits with reported receipt of UI payments during a calendar year. The first entry in each cell represents the frequency and the second entry denotes the percent of cases in each cell. To be deemed eligible a youth must have been entitled to receive at least one week of UI benefits at some time during a calendar year. Two sets of results are presented in the table corresponding to the two definitions of eligibility described above and in Section 4. The first set of 2 columns presents the results for our broad interpretation of the voluntary separation provision and the second set refer to the narrow definition of eligibility, which assumes all quitters are ineligible.

---

<sup>40</sup> A similar problem arises in the allocation of the number of weeks of unemployment to each calendar year when a spell overlaps two years. Again, the weeks of unemployment were assumed to occur at the beginning of the nonemployment spell. While this does not impact on the accuracy assessment, it does effect the calendar year measures of eligibility and utilization analyzed in Section 4.

TABLE B.4

Frequency Table of Imputed Eligibility and UI Receipt for Both  
Definitions of Eligibility  
(percentage of cases in each category in parentheses)

	Broad Definition		Narrow Definition	
	Ineligible	Eligible	Ineligible	Eligible
Nonrecipient	4214 (58.2)	1869 (25.8)	5134 (71.0)	949 (13.1)
UI Recipient	260 (3.6)	892 (12.3)	333 (4.6)	819 (11.3)

The results in Table B.4 are very encouraging. Using the broad definition of eligibility, an obvious error was made in only 3.6 percent of the cases: i.e., cases where a person reported receiving UI payments when we determined they were ineligible for benefits. A further examination of these 260 cases indicated that just 69 cases were judged to be ineligible because of insufficient earnings in covered employment, while the other 191 incorrect determinations resulted from the self-reported reason for beginning a nonemployment spell. Surprisingly, this type of error only occurs in 4.6 percent of the cases under the narrow interpretation of the voluntary separation provision. In addition, it is possible that erroneous eligibility imputations were made for the cases where we determined an individual was eligible for UI benefits but he or she did not report receipt of any payments. Alternatively, all 1869 or 949 cases, under the broad and narrow definitions respectively, could be the result of incomplete take-up rates for benefits.

Table B.5 presents summary statistics for the difference between reported average benefit payments and our imputed *WBA* for the years 1979 to 1984. Similar measures for the difference between weeks of receipt and the imputed value for weeks of eligibility are reported in Table B.6. In order to make the two measures used in this latter table comparable, the imputed measure for weeks of eligibility is set equal to the lesser of *WE* or the number of weeks of nonemployment during the year. The results in Tables B.5 and B.6 provide further evidence of the remarkable accuracy of our imputed measures of UI entitlements.

TABLE B.5

Summary Statistics for the Difference between Reported and Imputed  
WBA by Year for Broad Definition of Eligibility

Year	Lower Quartile	Median	Upper Quartile
1979	-3	1	19
1980	-8	1	17
1981	-11	1	23
1982	-11	1	21
1983	-19	0	21
1984	-11	2	29

TABLE B.6

Summary Statistics for the Difference between Reported Weeks of  
UI Receipt and Imputed WE Adjusted for Weeks of Nonemployment  
by Year for Broad Definition of Eligibility

Year	Lower Quartile	Median	Upper Quartile
1979	-3	-1	1
1980	-3	-1	3
1981	-4	-1	1
1982	-1	-2	2
1983	-2	1	8
1984	-3	-1	3

## References

- Atkinson, A. B. (1987), "Income Maintenance and Social Insurance", in A. J. Aurbach and M. Feldstein, eds., *Handbook of Public Economics*, Vol. 2, Amsterdam: North-Holland.
- Barron, J. M. and W. Mellow (1981), "Unemployment Insurance: The Recipients and Its Impact", *Southern Economic Journal*, 47, 606-616.
- Blank, R. and D. Card (1988), "Recent Trends in Insured and Uninsured Unemployment: Is There an Explanation?" Industrial Relations Section, Princeton University, Working Paper #243.
- Brechling, F. (1977), "The Incentive Effects of the U.S. Unemployment Insurance Tax", in R. Ehrenberg, ed., *Research in Labor Economics*, Vol. 1, Greenwich, Conn.: JAI Press.
- Burtless, G. 1983, "Why is Insured Unemployment So Low?" *Brookings Papers on Economic Activity*, No. 1, 225-249.
- Burtless, G. and D. H. Saks (1984), *The Decline of Insured Unemployment During the 1980's*. Final Report prepared for the U.S. Department of Labor, Employment and Training Administration, Office of Research and Evaluation.
- Cameron, A. C. and T. MaCurdy (1988), "The Extent of Measurement Error in Annual Earnings", Mimeo, Stanford University.
- Clark, K. B. and L. H. Summers (1979), "Labor Market Dynamics and Unemployment: A Reconsideration", *Brookings Papers on Economic Activity*, Vol. 1, 13-72.
- Clark, K. B. and L. H. Summers (1982a), "Unemployment Insurance and Labor Market Transitions", in M. N. Baily, ed., *Workers, Jobs, and Inflation*, Washington D. C.: The Brookings Institution.
- Clark, K. B. and L. H. Summers (1982b), "The Dynamics of Youth Unemployment", in R. Freeman and D. Wise, eds., *The Youth Labor Market Problem: Its Nature, Causes, and Consequences*, Chicago: University of Chicago Press.
- Classen, K. P. (1979), "Unemployment Insurance and Job Search", in S. A. Lippman and J. J. McCall, eds., *Studies in the Theory of Search*, Amsterdam: North Holland.
- Corson, W. and W. Nicholson (1988), *An Examination of Declining UI Claims During the 1980's*. Report prepared for the U.S. Department of Labor, Employment and Training Administration, Unemployment Insurance Service.
- Danziger, S., R. Haveman, and R. Plotnik (1981), "How Income Transfer Programs Affect Work, Savings, and the Income Distribution: A Critical Review", *Journal of Economic Literature*, 19, 975-1028.
- Ehrenberg, R. G. and R. L. Oaxaca (1976), "Unemployment Insurance, Duration of Unemployment and Subsequent Wage Gain", *American Economic Review*, 66, 754-766.
- Feldstein, M. S. (1978), "The Effect of Unemployment Insurance on Temporary Layoff Unemployment", *American Economic Review*, 68, 834-846.

- Feldstein, M. and D. T. Ellwood (1982), "Teenage Unemployment: What is the Problem?", in R. Freeman and D. Wise, eds., *The Youth Labor Market Problem: Its Nature, Causes, and Consequences*, Chicago: University of Chicago Press.
- Gustman, A. L. (1980), "Analyzing the Relation of Unemployment Insurance and Unemployment", in R. Ehrenberg ed. *Research in Labor Economics*, Volume 5, JAI Press Inc. 69-114.
- Ham, J. and S. Rea, Jr. (1987), "Unemployment Insurance and Male Unemployment Duration in Canada", *Journal of Labor Economics*, 5, 325-353.
- Hamermesh, D. S. (1977), *Jobless Pay and the Economy*, Baltimore: John Hopkins University Press.
- Hills, S. (1982), "Estimating the Relationship Between Unemployment Compensation and the Duration of Unemployment-The Problem of Nonfilers", *Journal of Human Resources*, 17, 460-470.
- Kaitz, H. B. (1970), "Analyzing the Length of Spells of Unemployment", *Monthly Labor Review*, 93, 11-20.
- Katz, L. (1986), "Layoffs, Recall and the Duration of Unemployment", National Bureau of Economic Research Working Paper No. 1825.
- Katz, L. and B. Meyer (1988a), "Unemployment Insurance, Recall Expectations and Unemployment Outcomes", National Bureau of Economic Research Working Paper No. 2594.
- Katz, L. and B. Meyer (1988b), "The Impact of the Potential Duration of Unemployment Benefits on the Duration of Unemployment", National Bureau of Economic Research Working Paper No. 2741.
- MaCurdy, T. (1985) "A Guide to Applying Time Series Models to Panel Data", Economic Research Paper, NORC.
- Meyer, B. (1988), "Unemployment Insurance and Unemployment Spells", National Bureau of Economic Research Working Paper No. 2546.
- Moffitt, R. (1985), "Unemployment Insurance and the Distribution of Unemployment Spells", *Journal of Econometrics*, 28, 85-101.
- Moffitt, R. and W. Nicholson (1982), "The Effect of Unemployment Insurance on Unemployment: The Case of Federal Supplemental Benefits", *Review of Economics and Statistics*, 64, 1-11.
- Newton, F. C. and H. Rosen (1979), "Unemployment Insurance, Income Taxation, and Duration of Unemployment: Evidence from Georgia", *Southern Economic Journal*, 45, 773-784.
- Topel, R. H. (1983), "On Layoffs and Unemployment Insurance", *American Economic Review*, 73, 541-559.
- Topel, R. H. (1985), "Unemployment and Unemployment Insurance", in R. Ehrenberg ed., *Research in Labor Economics*, Volume 7, JAI Press Inc. pp. 91-135.

- Topel, R. H. and F. Welch (1980), "Unemployment Insurance: What the Theory Predicts and What the Numbers (May) Show", *Economica*, 47, 351-380.
- U.S. Department of Labor, Employment and Training Administration, Unemployment Insurance Service, "Comparison of State Unemployment Insurance Laws", Various Issues.
- Welch, F. (1977), "What Have We Learned from Empirical Studies of Unemployment Insurance?", *Industrial and Labor Relations Review*, 30, 451-461.