

Survey of Income and Program Participation

Working Paper Series

No. 8404

Papers Presented in the
Survey of Income and Program Participation
Session II
at the Annual Meeting of the
American Statistical Association
in Philadelphia, PA
August 13-16, 1984

(#8404)

S.C.R.T.D. LIBRARY



U.S. Department of Commerce
Malcolm Baldrige, Secretary
Clarence J. Brown, Deputy Secretary
Sidney Jones, Under Secretary for
Economic Affairs

BUREAU OF THE CENSUS
John G. Keane,
Director

06062

HA
12
.A43
1984
v.1



U.S. BUREAU OF THE CENSUS

John G. Keane, Director
C.L. Kincannon, Deputy Director
William P. Butz, Associate Director
for Demographic Fields

POPULATION DIVISION

Roger A. Herriot, Chief

Acknowledgements

This publication is composed of papers prepared by many different authors for presentation at the American Statistical Association on August 13-16, 1984. We would like to thank these authors for their cooperation in making the papers available for publication. Clerical and editorial assistance was provided by Hazel Beaton, Mary Kisner, and Delma Frankel.

Suggested Citation

"Papers Presented at the Survey of Income and Program Participation Session II at the Annual Meeting of the American Statistical Association in Philadelphia, PA, August 13-16, 1984," SIPP Working Paper Series No. 8404. U.S. Bureau of the Census, Washington, D.C. 1984.

Preface

This report is comprised of five papers featured in the "Survey of Income and Program Participation" Session II, one of three in the Social Statistics Section of the Annual Meeting of the American Statistical Association.

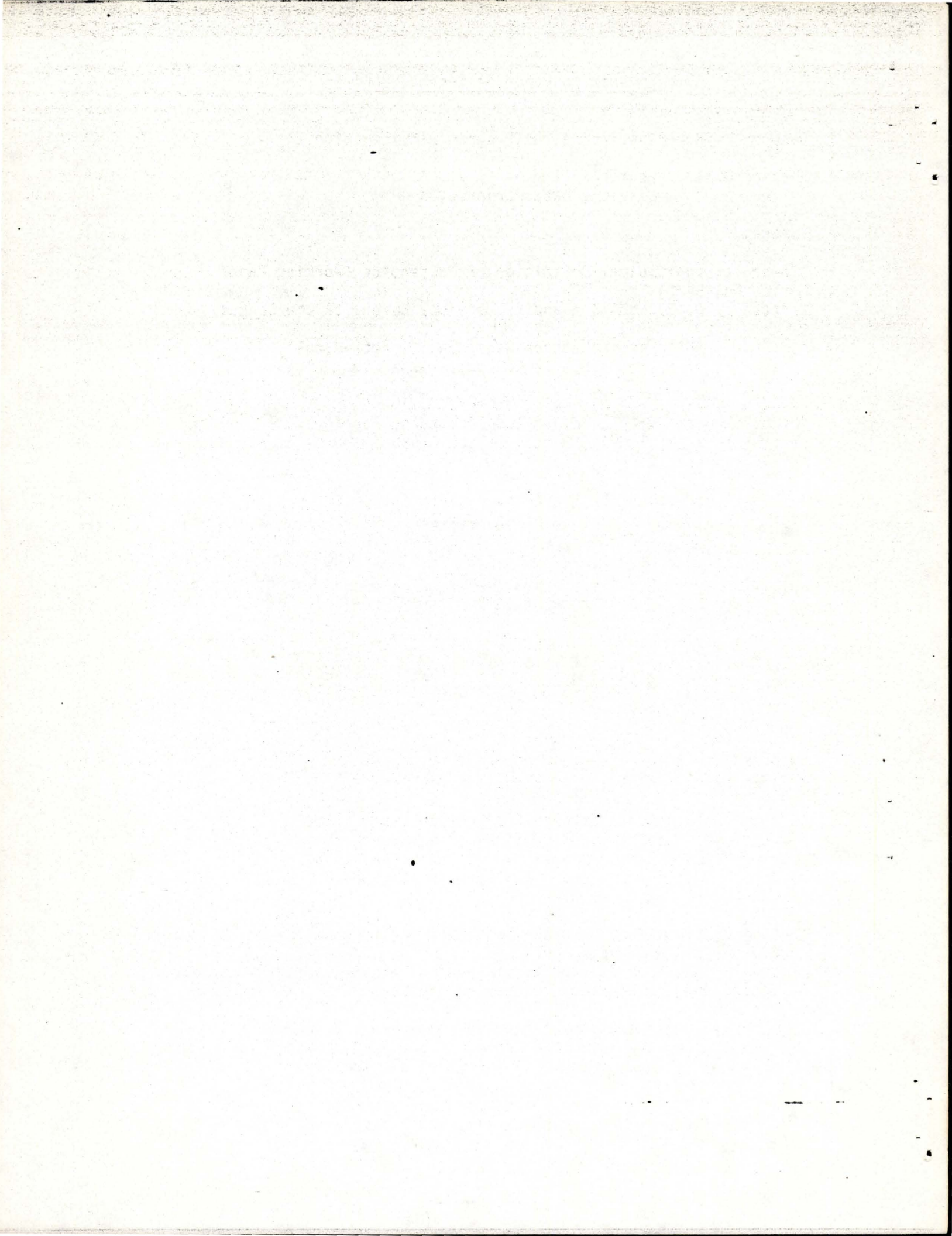
These sessions covered a range of topics, both methodological and substantive, about longitudinal surveys and the Survey of Income and Program Participation (SIPP).

SIPP is a new Census Bureau survey collecting data that will help measure income distribution and poverty throughout the country more accurately. These data will be used to study Federal and state aid programs (such as food stamps, welfare, Medicaid, and subsidized housing), to estimate future program costs and coverage, and to assess the effects of proposed changes in program eligibility rules or benefit levels.

Households in the survey will be interviewed at 4-month intervals over a period of 2 1/2 years. The reference period will be the 4 months preceding the interview. In all, about 20,000 households will be interviewed, approximately 5,000 each month. Field operations will be handled through our 12 regional offices.

Recurring questions will deal with employment, types of income, and noncash benefits. Periodic questions will be added dealing with school enrollment, marital history, migration, disability, and other topics. Special supplemental questions will also be added to the SIPP questionnaire.

These papers discuss SIPP and its predecessor, the Income Survey Development Program (ISDP), an experimental program designed to test procedures used in conducting SIPP.



Working Paper Series No. 8404

Session: Survey of Income and Program Participation II

Session Chair: George Hall
Baseline Data Corporation

Papers:

"Toward a Longitudinal Definition of Households (Working Paper Series No. 8402)."

Written by D. B. McMillen and R. Herriot, Bureau of the Census.

(Discusses the cross-sectional/longitudinal conflict and attempts some reconciliation.)

"Lifetime Work Experience and Its Effects on Earnings."

Written by J. M. McNeil, Bureau of the Census, and J. T. Salvo, New York City Department of Planning.

(Discusses differences between men's and women's work experience and the impact on earnings.)

"Panel Surveys as a Source of Migration Data."

Written by D. Dahmann, Bureau of the Census.

(Discusses analyses of migration and previous panel surveys to assess how SIPP can further our understanding of geographical mobility processes.)

"SIPP and CPS Labor Force Concepts: A Comparison."

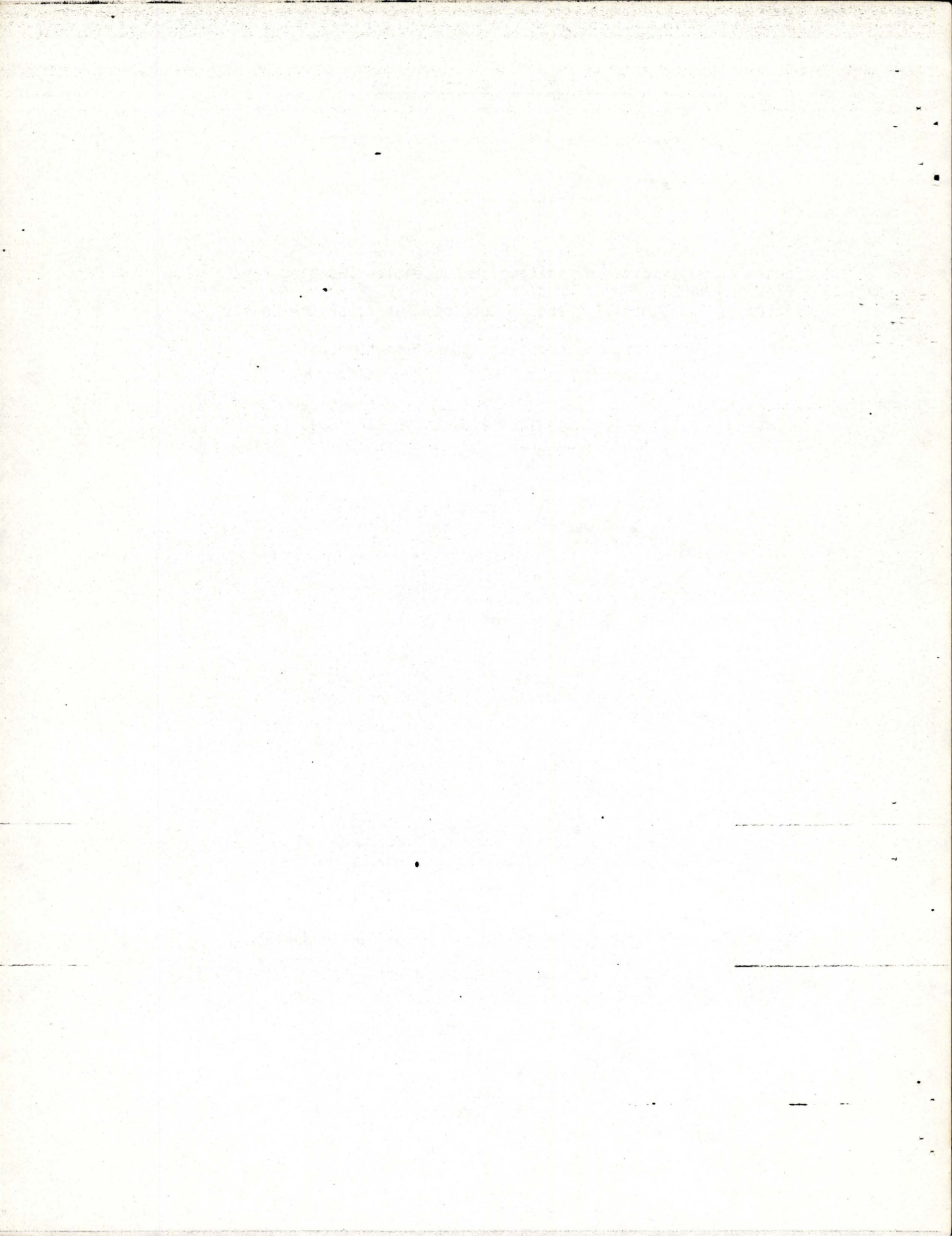
Written by P. M. Ryscavage, Bureau of the Census.

(Discusses differences between labor force concepts found in SIPP and the Current Population Survey, our official source of employment and unemployment estimates.)

"Matching Economic Data to the Survey of Income and Program Participation: a pilot study."

Written by S. Haber, George Washington University, and P. Ryscavage, D. Sater, and V. Valdisera, Bureau of the Census.

(Discusses matching data from SIPP with data from files such as the Standard Statistical Establishment List, which contains employer data on payroll and sales receipts. Focuses on the mechanics of the matching process and the uses of the data.)



TOWARD A LONGITUDINAL DEFINITION OF HOUSEHOLDS

**David Byron McMillen
and
Roger A. Herriot**

**Population Division
Bureau of the Census
Washington, D.C. 20233**

**For presentation at the annual meeting
of the American Statistical Association,
Philadelphia, Pa., August 13-16, 1984.**



U.S. BUREAU OF THE CENSUS

John G. Keane, Director

C.L. Kincannon, Deputy Director

**William P. Butz, Associate Director
for Demographic Fields**

POPULATION DIVISION

Roger A. Herriot, Chief

Acknowledgments

This paper was prepared by David Byron McMillen and Roger Herriot, Population Division. This paper has benefited from the comments of a number of our colleagues at the Census Bureau. Special thanks go to Daniel Kasprzyk and Paul Siegel who have read and commented on innumerable drafts.

Suggested Citation

McMillen, David Byron, and Herriot, Roger. "Toward a Longitudinal Definition of Households." SIPP Working Paper Series No. 8402. U.S. Bureau of the Census, Washington, D.C. 1984.

Introduction and Background

Data collection and analysis in the social sciences generally focus on cross-sectional surveys such as the Current Population Survey (CPS). Consequently, most of our concepts and data analysis tools are structured around point estimates of some phenomenon or characteristic. To the extent that we try to develop longitudinal concepts and measures of social phenomena, that is to say viewing events across time rather than at one point in time, we conflict with these cross-sectional structures. It is the goal of this paper to confront that cross-sectional/longitudinal conflict and attempt some reconciliation. More specifically, this paper attempts to develop longitudinal definitions of households and families which are useful for observing these units across time and for constructing aggregate characteristics across that time period, while not creating serious conflict with our cross-sectional constructs of household type and composition. We begin this exercise by examining cross-sectional household concepts from the CPS and recounting the deficiencies of that perspective. Next we will examine several types of longitudinal definitions, identify the type that is usually cited as most useful, and describe three definitions within that framework. In the third section of this paper we will evaluate the definition options available, in terms of utility as well as what is possible given the data at hand to implement such a definition. Next we will illustrate how this definition might be used in calculating aggregate household characteristics and in tabulations of the number of households, household types, and household characteristics.

Point Estimates and Longitudinal Measures

The household definitions used in the CPS serve as adequate measures for the intended cross-sectional purposes. Indeed, few people argue that these definitions create a problem when estimating the number of households by type at the time of the survey. However, when those definitions are used in conjunction with other variables, problems begin to develop.

Discussions on measuring annual household income from the CPS center around the retrospective nature of the measure. Household members as of March are asked to recall their income for the previous calendar year, and the income of all members are aggregated to create a household income. The problem centers on the varying lengths of household membership and the unvarying time period used in the aggregation. Even though a person is a member of the household for part of the year and, thus, contributes income for only part of the year, that person's entire annual income is included in the household income.

Similarly, persons who are not members of the household at the time of the survey are not included in calculating the annual household income, even though they may have contributed income for most of the year. This type of criticism is often used to question the adequacy of the CPS income measures; however, it is better viewed as an example of the problems created by combining a point estimate of household composition with a longitudinal (annual) measure of income. Inevitably, the compromises necessary to combine such cross-sectional and longitudinal constructs produce a less than ideal measure.

Similar criticism of the CPS household data can be made. If we examine consecutive March measures of the distribution of households by type we observe little change. The CPS measure of households masks most of the interesting change in the distribution of households. For example, in recent years the number of married couple households has changed at a rate of less than 1 percent a year, or about 200,000 households. Concurrent with that indistinguishable change are over 3 million marriages and divorces, not to mention changes in household type as a result of the death of one member. The small net change creates the appearance of stability, while masking considerable activity. Again, the problem is not so much the inadequacy of the data, but rather the difficulty of measuring longitudinal events with point estimates.

When criticism is leveled against a particular measure, the problem often is not the measure but rather the incongruity between the measuring instrument and the time frame being considered. The examples used above are annual measures, but the same problem exists regardless of the length of time. Most social measurement is discrete while time is continuous. The goal of course is to get to the point where the difference between the two is trivial and can be easily ignored.

In summary, much of the criticism of CPS measures can be attributed to this discrepancy between the time reference of the social measurement and the cross-sectional survey instrument. One solution to the problem is to decrease that difference by repeatedly measuring the phenomenon in question during the year. Those observations can then be aggregated to produce measures which cover a number of time intervals. It is from this perspective that the design for the Survey of Income and Program Participation (SIPP) has developed.

The design of this survey is to interview the household every 4 months over a 2 1/2-year period, and to collect in those interviews monthly data on household composition, income, labor force participation, and a number of other characteristics. Those monthly data can then be aggregated to larger temporal units such as quarterly or annual measures. However, with the idea of aggregating monthly units comes the problem of defining which units should be aggregated across time and which should not. That is to say, which households are the same over the period, which exist at the beginning of the period but not at the end, and which exist at the end but not at the beginning. Without such a definition, aggregating above the person level is impossible.

Defining households across time is an issue that has been debated for several years without resolve; however, it is necessary that the Census Bureau decide

which of the many proposed methods will be used for the publication series from the Survey of Income and Program Participation (SIPP). This paper moves one step nearer to that decision by summarizing the proposals on how longitudinal households should be defined and recommending a system to be used. In addition, this paper will begin to identify what conceptual and processing problems remain unsolved given the definition chosen.

Several proposals have been offered for defining longitudinal households. Griffith (1978) outlines six measures, one of which is the traditional Current Population Survey (CPS) definition. Griffith also proposes that the Census Bureau use several definitions in tabulating households from SIPP. Others who had proposed variant measures include two from Davey (1980), Crosby (1979), Lane (1978), and Smith-Ycas (1981). Ycas (1981) in summarizing past work identifies four keys for labeling definitional methods: static; dynamic; static-dynamic hybrids; and attribute methods. In the following section we will discuss several types of longitudinal definitions for households.

Types of Longitudinal Household Definitions

Static definitions of households fix the household composition and characteristics at a given point in time and calculate other attributes from that point. These definitions are the standard cross-sectional perspective on households common to the CPS and other similar surveys. Using a point estimate of household composition, other attributes are calculated assuming that the composition chosen existed for the full period. Thus, some estimate of annual income for each member is aggregated to produce an annual household income, regardless of whether members were there for the full period or joined the day before the interview. This type of household definition is the logical outgrowth of cross-sectional surveys where interviews are conducted at one point in time

and aggregates of past events are a function of respondent recall. This type of definition coincides with the instantaneous conception of a household which we use from day to day.

Static definitions are both useful and familiar for cross-sectional surveys; however, they serve little purpose in longitudinal surveys other than to provide familiarity. Static definitions, for a number of reasons, ignore the dynamic activity common to households--households are formed by marriages and dissolved by divorces, children leave home and set up their own household, or move in with relatives, and so on. It is difficult to justify the expense and complexities necessary to measure these dynamics if we then suggest to ignore them in defining households. It is useful to portray static definitions here, however, for they represent one end of the definition spectrum.

Dynamic definitions of households occupy the other end of the spectrum. These definitions recognize change as inherent in observing households across time, and attempt to incorporate that change into the definition. Thus, household characteristics and attributes become variables to be measured as households change, are created, and dissolve during the period of observation. In other words, these definitions attempt to minimize the extent to which dynamic concepts are forced into static categories. Needless to say, dynamic definitions are better suited to a longitudinal survey such as SIPP; however, such definitions are difficult to devise and to carry out.

One of the first difficulties encountered with dynamic definitions is that they produce measures which are not readily familiar to many of those who use census data. The most common illustration of this point uses household size. Static definitions produce measures of household size such as 2 or 3, which are intuitively meaningful. That is, they fit with our instantaneous image of house-

holds because they represent the household size at one point in time--the survey date. Dynamic definitions produce measures of household size which look more like averages across a number of households--2.4 or 3.2 members in the household. These measures are summary statistics of the household experience, summarizing across time. In other words, dynamic definitions force us out of that instantaneous view of households and into thinking about them as something which change across time; our statistics produce a summary of that change. Ycas (1981) suggests several ways of handling the problem of household size-rounding, using modal size, etc.--however, it may be best to reeducate the reader to think of annual household characteristics as the aggregate of a number of discrete experiences. There are other more troublesome problems to be dealt with in developing dynamic household definitions. I will deal here only with definitional problems, acknowledging that there are also measurement problems to be considered later in this paper.

Unlike many demographic variables, there are several aspects of dynamic households for which there is no definition or consensus as to what constitutes a change in type. For example, if a husband and wife divorce, there are several ways we can account for this on our household ledger. We could count this as the dissolution of the husband/wife household and the formation of two new households. This results in a net increase of one household in existence at that point in time and an increase of two households when counting the number of households existing during the period. Alternatively, we could allow one household to be the continuation of the husband/wife household. Again we have a net increase of one household in existence, but because of the continuing household, we increase only by one the number of households existing during the period. To generalize, a household may experience a number of changes across time and we can converse easily about the discrete events. However, we do not have a clear

concept of when those changes result in the formation of a new household and the dissolution of an old household. One extreme is to say that any change to the household composition results in a new household. At the other extreme are those who say that this is an issue without resolution and suggest that we abandon the measurement of household characteristics except as they pertain to individuals. In other words, before we can implement a dynamic definition of households, we must first develop a set of continuity rules or accounting principles which identify cases of household dissolution, household formation, and cases where two households at two points in time are identified as the same household.

Most longitudinal household definitions that have been proposed fall somewhere between the static and dynamic extremes. Each acknowledges the difficulty of developing continuity rules, and proposes some static-dynamic blend to finesse those problems. A number of cross-sectional/dynamic hybrid definitions have been proposed. One set of these definitions is quasi-dynamic and acknowledges that a set of continuity rules has yet to be developed. Another set is basically a static system designed to avoid the continuity dilemma. Neither of these alternatives is particularly attractive. In the latter case, most of the alternatives create as many problems as they solve. In the former case, if we are going to develop a set of continuity rules, then there is little need for a hybrid definition.

Attribute-type definitions are drawn from the work done on the Panel Study of Income Dynamics (PSID). The goal in these definitions is somewhat different than in the previous discussion. Rather than attempting a longitudinal definition of households, this system calculates a series of cross-sectional households at some smaller time interval, and then ascribes the characteristics of

the household to each individual. Measures for some longer time interval are then calculated by aggregating the series of point estimates across each individual to represent that person's household experience during the period. This system will yield the number of persons who lived in "households" with a monthly income of \$1000 to \$1500 during the year; however, it is more difficult to derive the number of households with an annual household income of \$12,000 to \$18,000. In fact, without an additional set of assumptions, this system does not produce an accounting of households across time. In order to develop household statistics within the attribute system it is necessary to assume, for example, that the householder at the end of the period will represent the household experience. Then the household attributes ascribed to that person are aggregated to produce household characteristics. Those aggregated attributes represent the householder's experience during the period, but not necessarily the experience of the other persons in the household at that time. As can be seen, an assumption such as this contains many of the weaknesses of using a static or cross-sectional definition of households, with few obvious advantages at the household level.

In summary, there are four types of household definitions which have been proposed for use with longitudinal surveys: cross-sectional; dynamic; cross-sectional/dynamic hybrids; and the attribute system. The cross-sectional approach is clearly inappropriate since it ignores the dynamic nature of the data. The attribute system incorporates the dynamic aspects of the data but dodges the issue of developing continuity rules for households. Consequently, this system, by ignoring the social structure of households, produces many of the same problems raised in criticism of the CPS measure of annual household income. It is clear that a dynamic definition is the most desirable alternative, but agreement on just how that definition ought to be formed is elusive.

hold. He proposes that we acknowledge change between the following types of households: 1) married-couple household; 2) male family-household; 3) female family-household; 4) male nonfamily-household; and 5) female nonfamily-household. Thus any change which results in a household falling into a different category results in the dissolution of one household and the formation of another. To illustrate this system consider a husband-wife household which experiences a divorce. Norton's scheme would consider the husband/wife household dissolved and two new households formed. The two new households would be family households if there were children present and nonfamily households otherwise.

A second longitudinal definition has been proposed by Dicker and Casady (1982) for use with the National Medical Care Utilization and Expenditure Survey (NMCUES). In a slight departure from the definitions discussed here, Dicker and Casady focus on families rather than households; however, that does not pose any serious problems. Like others who have approached this problem, Dicker and Casady begin with the realization that there is not consensus on when families begin or cease to exist; rather, such transitions are in part a function of the problem being investigated.

The NMCUES model for defining longitudinal families requires that antecedents and descendent families or, in their terms, predecessor and successor families be defined reciprocally. That is to say, any rules defining relationships across time must be applied to both families simultaneously. Dicker and Casady next demonstrate that when applying these rules you wind up with links between a number of households. That is to say, any family is likely to have more than one predecessor and more than one successor family. Thus, the problem lies in defining which of the possible pairs will be defined as the longitudinal

family. As with most longitudinal definitions, the system eventually reduces the decision to what will be defined as the same and what will be defined as change.

Dicker and Casady chose to define sameness by a majority rule. The successor family which receives the majority of members from the predecessor family is identified as the "principal predecessor." These two families then form the linked or longitudinal family unit. Finally, in cases where families split evenly, it should be noted that the NMCUES model does not define a longitudinal unit, but rather dissolves the predecessor family and considers all successor families as newly formed.

Five rules of relationships focus Siegel's (1982) development of household demography. The first two state that a household can have only one descendent and one antecedent that is identified as the same household. That is to say, when a household splits, only one of the subsequent households can be identified as "the same" household. The third and fourth rules identify households which are not the same as some preceding or succeeding household. Households which are not the same as any antecedent household are newly formed; a household not the same as any descendent household is dissolved. The final rule, one of transitivity, states that if A is the same as B and B is the same as C, then A must be the same as C. All that remains to complete this set of accounting principles is a definition of sameness. The rule is offered that two households separate in time and having the same householder are the same household.

Continuity based on following the householder has been criticized because of the somewhat arbitrary way in which the householder is defined, and because it creates what some consider unreasonable change within a continuing household. The most frequently cited example of such change is following the male after a

divorce when the children remain with the female. An alternative to Siegel's householder rule which is consistent with his demography of households is to follow the principal person. The principal person is the female in the married-couple household and the householder in all other households: this is the concept used in developing household weights in CPS. By following the principal person, we alleviate the problem cited above. Of course the problem now occurs when the children stay with the male following a divorce, a much less frequent event.

FIGURE 1

	<u>Norton</u>	<u>Siegel</u>	<u>Dicker</u>
Time 1	ABcd*	ABcd*	ABcd*
Time 2	A* Bcd*	A Bcd*	A* Bcd
Time 3	Ac* Bd	Ac Bd	Ac Bd
Time 4	Ac d B*	Ac d B	Ac d B

*New household

Let us consider briefly the strengths and weaknesses of these three systems focusing on two issues: 1) the number of households created across time, and 2) the extent to which the definition promotes or discourages longitudinal analysis. Norton's system comes the closest to maximizing change and, as a result, creates more households than the others. Consider the divorce example from above, but two children remain with the female. Both Siegel and Dicker and Casady would produce a total of two households resulting from the divorce. Norton's system produces three households: 1) the original married-couple household; 2) a male nonfamily-household; and 3) a female family-household. Let us continue following these people and assume that the children leave the female one at a time and join the male (see figure 1). In Norton's scheme, the first move by a child would produce the dissolution of the male nonfamily-household and the creation of a male family-household. Our longitudinal count

of households now stands at four. Neither Dicker and Casady nor Siegel would produce new households as a result of the children moving. When the second child moves, the female family-household is dissolved and a female nonfamily-household is created. The male family-household remains unchanged. Over these four observations, Norton's system produces five households; both Siegel and Dicker and Casady produce only two by allowing the continuation of a household across these changes. Let us then look at those continuing households. The continuous household for Siegel's householder rule starts as a four-member married-couple household, dwindles to one member--the male, and increases to two with the addition of one child and then to three with the addition of the second child. On the other hand, Dicker and Casady's continuous household begins as the four-person married-couple household and is transformed by the divorce to a three-person female-headed household, then to a two-person and, subsequently, a one-person nonfamily-household. It should be noted that these two continuous households follow opposite courses after the divorce. The continuous household under the principal-person rule would be identical to Dicker and Casady's continuous household.

We should stop at this point and examine what we can learn from this cumbersome yet realistic example of household change. First, if we are interested in counting households (the number or percent of households in poverty during the year, for example), then a continuity system such as Norton's, which allows for continuity in only the most trivial cases, creates a much larger number of households. Suppose the female half of our mythical household was in poverty after the divorce. By Norton's count, during that year we would have 20 percent of the households in poverty. Dicker and Casady and Siegel would show 50 percent of their households in poverty. A second observation to be made

here is that, regardless of what sort of continuity rule we adopt, we will observe households which contain a wide variety of change. The question we must ask is whether we accept that households undergo such change and remain intact.

As noted above, each of the three systems has its constituency and its detractors. Siegel's system is criticized because of the disjunctures that can occur following a divorce. For example, the continuous household will follow a male householder who divorces his wife even though the wife and children remain in the housing unit as a group. Similarly, Norton's scheme is criticized because of the lack of attention paid to continuity. Dicker and Casady are criticized by the mechanical nature of the majority rule. Why, it is asked, should one person make all the difference in whether a family is designated new or continuous?

None of the definitions of longitudinal households offered in the literature has proved viable. However, in the process of discussing this issue with several demographers and economists, it became clear that any definition which labels a transition from a family household to a nonfamily household as continuous causes problems. Drawing on that experience, we determined that we should develop a longitudinal definition of families separate from that for nonfamily households.

We begin with the CPS definition of a family as two or more persons, one of whom is the householder, related by birth, marriage, or adoption, and residing together. To make this cross-sectional definition dynamic, we must add the time dimension or develop a continuity rule. Thus, a longitudinal family is defined as two or more related persons, at least one of whom is the householder

or spouse of the householder, who had the same household experience over two or more consecutive months. We further stipulate that no more than one core family unit with children can continue from a previous-month family. Three levels of criteria are offered to distinguish cases where both parents and children split into two or more households. The first-level criterion, for continuity, is that the family with the most child-months is identified as continuous. The second level, to distinguish between families with the same number of child-months, is the family with the most family-months. In the third level, if two potential continuing units tie on both of the above criteria, then the continuing unit will be assigned randomly. Two elements have been added to the CPS definition: 1) the time dimension, and 2) the inclusion of spouse as part of the continuity criteria.

Let us examine this definition more carefully. Consider again our four-step example of divorce and then the movement of two children one at a time from one parent to another. Following the separation, the Bcd family would be the continuing family because it contains two or more members of the initial family, one of whom is the householder or spouse. The A household would be new because of the transition from family to nonfamily status. Following the movement of the first child, c, the Ac family is considered newly formed because of the transition from nonfamily to family status. Finally, the movement of d from the Bd family to the Ac family results in a new nonfamily household, B. Using the notation from figure 1, we have:

- | | | |
|-----------|--------|----|
| 1. ABcd* | 3. Ac* | Bd |
| 2. A* Bcd | 4. Acd | B* |

Next, we must confront defining continuity for nonfamily households. A non-family household is two or more unrelated individuals residing together or an

individual living alone. For these cases, we have adopted a 50-percent rule. As long as the householder and 50 percent or more of the household is the same at two points in time, the household is considered a continuous household. The distinction between this and the majority rule is that, rather than creating new households for even splits, this rule provides for continuity.

Given the above definitions for longitudinal family and nonfamily household, we have now defined the full set of longitudinal households in SIPP. However, other possible longitudinal units or groups exist in relation to federally funded support programs. For example, food-stamp units and AFDC units are defined independently of the household and, in fact, households may contain more than one of these units. Longitudinal units for these programs will be defined on the basis of the person in whose name the program application is filed. For example, in a husband-wife, two-child family, the male is defined as the food-stamp recipient. If he leaves, that food-stamp unit is dissolved; a new one is formed if the female reapplies and is found eligible.

Perspectives on Household Characteristics

In this section, I will address the uses of this longitudinal definition of households and argue that we need to tabulate household data from SIPP using at least two types of longitudinal definitions. The need for two types of definitions is a function of the kind of household information needed. Before I develop this argument, it is useful to record here an analogy developed by Paul Siegel to illustrate the function of a longitudinal definition.

Consider, if you will, a chemistry research lab with a number of rooms in which radioactive elements are used. This lab also has a number of employees who work with these radioactive materials, and, like any good bureaucracy, the workers find themselves constantly on the move. Each week, when they come to work, their

workroom assignment has changed and they go off to work with yet another radioactive element in yet another room. To complete our analogy, consider the rooms in this lab to be our households and the workers the members of those households. Fortunately, the rooms neither move nor change shape, so sameness isn't a problem here, but there are other aspects which are helpful. As good administrators, we are worried both about the safety of our researchers and staying within the guidelines for the maximum allowable accumulation of radioactivity in a lab. And so we set out to measure radiation levels when it dawned on us that we must take two measurements. First, we must determine the radiation level in each room; since that varies, as does the amount of time each worker spends in each room, that is not enough. To measure the exposure of each worker, we also must measure the time spent by each worker in each room. Our analogy has a fairytale ending; in this case, we were able to rely on administrative records to collect the room-duration data, and our study came in on time and under budget.

A number of useful parallels can be spun from this analogy, but I will confine myself here to a discussion of how we will look at these data. First, our households/rooms are our longitudinal unit, and part of what we are interested in is characteristics of that unit. How many people have been members, or what is the membership of that unit over the year? We also want to characterize the room by the amount of radioactivity (income) collected there and the changes that have occurred within that unit during the year. However, these tables do not tell us much about the well-being of our workers. Thus, we also need to examine individual experiences. What types of rooms did they work in during the year? How can we characterize the accumulated experiences in those rooms?

Going back to our types of longitudinal definitions, I am arguing that we need to use both a dynamic longitudinal and an attribute-type household definition, because we are interested in both the experience of households and of individuals in households. There are some characteristics like annual household income which we need to examine both as a characteristic of the household and of the individual. This is only to say that there are multiple meanings associated with the concept of income. In CPS, where we have only one way of obtaining income data, we attach all of those meanings to that single measure. SIPP allows us to decompose that measure and look at the components more carefully.

To summarize, I have argued that to fully appreciate the household dynamics we observe in SIPP and to portray that activity over a year, we should provide two types of tabulations. The first tabulates household characteristics using a longitudinal definition and examines how changes in some characteristics result in changes in others. The second type of tabulation examines how household characteristics affect individuals across time.

It is useful at this point to make one other comment about dynamic household tables. Regardless of the definition used, for any given period of time, we will have one set of households which existed for the full period and one set which did not exist for the full time period. For a number of reasons, we should tabulate these groups separately. The first group has a full set of information and present no problem. The second group, those households newly formed or dissolved during the year, are quite different. First, they do not have data for the full year. Second, we know that they have experienced some change--how drastic depends on the definition chosen--and, thus, we are likely to want to examine different items for these cases. Finally, we cannot calculate annual characteristics for part-year households without imposing a set of assumptions.

Calculating Household Income

Regardless of the definition chosen, many households will be continuous yet experience some demographic change. It then must be specified whose income to include for what periods of time. Thus, even though we define most households as the same, those households are likely to experience some compositional change. It is that compositional change that complicates the calculation of an annual household income. For households that remain the same but experience some compositional change, we propose that income be aggregated from the monthly responses across all members of the household for the duration of the membership for the full year. Thus, if two persons were members of a household for the full year and a third person with income joined for 3 months, a total of 27 person-months of income would be summed to produce an annual household income. We propose a similarly straight-forward method for calculating the average household size. The household size is the sum of person-months divided by the number of months in the period (12 for years; 3 for quarters, etc.). In the case above, the annual household size would be 2.25. This procedure for calculating household income has a number of advantages. First, it accumulates income only for those persons in the household and thus avoids the inflation/deflation problems inherent in a static definition. It also maintains the short recall advantages built into SIPP. Other advantages include the conceptual clarity of the definition, as well as some processing advantages which will be discussed later.

While the calculation of annual household incomes for households which remain the same is rather simple, some problems remain with defining such income for households which exist for only part of the observation period. Using the procedure above, we can calculate aggregate annual household income for households that are born or die during the period; however, it is not clear that

these incomes should be treated as the same information as that calculated for households which existed for the full year. To illustrate some of the problems, let us take a household with one member there for 8 months with an income of \$500 per month, and a second member present for 4 months with an income of \$600 per month. The aggregate income for that household is \$6400; however, the annual income is somewhat between that minimum and \$13,200 which assumes that both household members earn at that rate for a full 12 months. An additional problem is raised in the calculation of household size. For this example, the household would have a annual size of 1 (12 person-months/12) unless we propose some other rule for defining the denominator. Using the greatest household duration for all members would yield a household size of 1.5 (12 person-months/8); however, it would not be comparable to an annual household size. The income problem is somewhat less problematic for households which dissolve than for households that are formed. Dissolved households will earn no more income, and, thus, we have complete information for that household. Newly formed households are likely to have their income truncated because they did not exist at both time points being measured. One solution for both the income and size problems is to tabulate separately those households that remain the same and those households newly formed or dissolved during the observation period. Some alternate measure from annual income, such as average monthly income, or an annualized income could then be computed for those households. Similarly, household size could be calculated using the number of months the householder was present in the denominator and person-months in the numerator.

Another option for tabulating part-year households is to weight them by their duration. Thus, a household existing for 6 months is counted as one-half of a household; a household existing for 9 months is counted as three-fourths of a household; and so forth. However, it is still an issue as to which cell of the

annual income these households should be tabulated. To tabulate a one-half-year household with 6 months of income totalling to \$6000 as a household with a \$12,000 annual income, a constant flow of income would have to be assumed. Similarly, to tabulate that household as one-half a household with an annual income of \$6000 is also misleading. A more accurate representation of this situation is to tabulate it as one-half a household with an average monthly income of \$1000. This incorporates both the time dimension of the household and the recorded income information without distorting the tabulation of income. If, on the other hand, what you want is a tabulation of annual income, the solution should be to tabulate only those households that are defined to exist as a continuing unit over the full period and to clearly note that this represents a limited or restricted universe of households. If this option is pursued, some research should be initiated to investigate whether any biases are introduced by censoring the data in this manner.

It should be noted here that, while annualization is mentioned as an option, it is included only because it is an option many have proposed. I contend that annualization is the worst possible solution. Our goal is to measure the income received by a household over a given period of time. Much of the effort of SIPP is designed to give as accurately as possible a picture of the changing membership of those households across time. To compute an annualized income for a household that is newly formed during the period, it would have to be assumed that the household existed unchanged across that period; we know that is not true. As noted above, we must live with the dilemma that for some households it is not possible to compute annual aggregates. It is possible, however, to compute subannual measures for nearly all households (at least for households that exist for one month or more). For example, the monthly rate of income can be computed for all households and is a reasonable base of ---

comparison. However, to compare an income flow measure with an annual income stock creates more problems than it solves.

Longitudinal Household Weighting

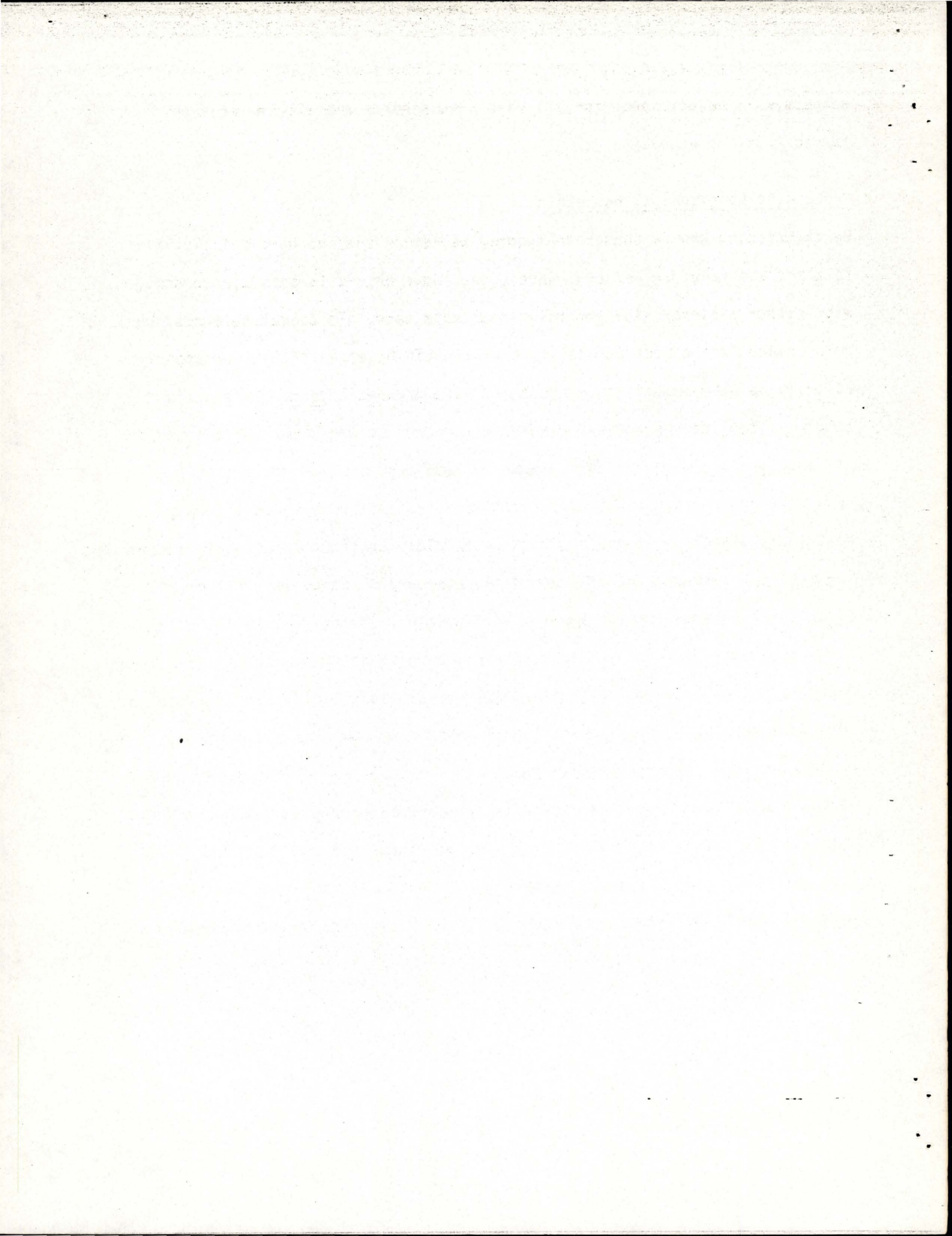
Having developed a longitudinal household definition and having begun to lay out the ways in which we want to aggregate and tabulate household data, it is now time to confront the issue of developing longitudinal weights for the households in SIPP. The target population of households used for these weights is all households that exist during the 2 1/2 years covered by the SIPP panel. Our longitudinal definition allows us to identify all households in existence at the beginning of the survey and continuing throughout, as well as all households that are newly formed from those households or dissolved. All of these households can be linked to the population from which the initial sample was drawn. There is, however, a set of households which are formed from persons outside the sample. Anyone residing overseas, in an institution, or in an armed forces barracks at the time of the first interview who subsequently leaves that special population will not be captured by this survey until they come in contact with a person or household in the sample. Any households formed by persons from one of those special populations will not be captured by this survey. Without an adjustment for the omission of these newly formed households, we will consistently underestimate the number of households after Wave 1.

Drawing on the experience of the National Center for Health Statistics (NCHS), it is suggested that weights be developed only for households containing a person from the Wave 1 sample at the beginning of the survey. Weighting of other households would require an extensive set of assumptions. In other words,

weights will be developed for all Wave 1 households and all new households formed from those households.

Processing Longitudinal Households

By calculating annual household income, we assume that we have a fully linked file for the calendar period in question. However, it is possible to proceed with either a hierarchical or relational data base. To calculate annual household income from a hierarchical data base would be easiest if all persons who at any time were members of a household were grouped together. This could result in duplicate records for some persons, or it would mean that antecedent and descendent households would need to be contiguous. In addition, each person record would have to carry some set of identifiers indicating what months that person belonged to which household. In other words, before computing the annual income measures, it would be necessary first to link all records so that we had information for the first five waves. Second, households defined as the same, newly formed, or dissolved must be identified and the links between newly formed and antecedent households must be identified. Third, the data set should be sorted so that each same household is followed by its descendent households. Thus, when processing for annual income, we will know that, when we begin to process a household which is unrelated to the previous household, we have processed all cases for that previous household and all of its descendents. Finally, separate tallies can then be made for full year annual income and for annualized income for newly formed and dissolved households.



REFERENCES

- J.E. Griffith. Impact of a Longitudinal Survey on the Data Base. (ISDP Working Paper #1, 1978.)
- E. Davey. Memorandum for the record. 12/24/80.
- B. Crosby. Memorandum to Lane, Ycas, and Mahoney. 7/79.
- J. P. Lane. Four Alternative Ways of Defining "Annual" Analysis Units. (ISDP Working Paper #6, 1978.)
- Smith-Ycas. Developed by Ycas based on reports of proposal by J. Smith. 1981.
- M. A. Ycas. Income Analysis Units for Longitudinal Files. 1981.
- A. J. Norton. Notes prepared for 11/18/82 meeting of SIPP Household Task Force.
- P. Siegel. Notes prepared for the 10/81 meeting of the Census Advisory Committee on Population Statistics.
- M. Dicker and R. J. Casady. A Reciprocal Rule Model for A Longitudinal Family Unit. Paper presented at the 1982 meetings of the American Statistical Association.

PRELIMINARY 7/18/84

Lifetime Work Experience and Its Effects
on Earnings: Data From the Income Survey
Development Program

John M. McNeil
Population Division
U.S. Bureau of the Census

and

Joseph J. Salvo
N.Y. Department of City Planning

For presentation at ASA Annual Meeting, Philadelphia, August 1984.

The extent to which persons remain attached to the labor force over the course of their working-age years has important economic and social implications. Differences in labor force attachment between men and women has been cited as one major reason why women earn less than men. This study presents data from the 1979 Income Survey Development Program (ISDP) on lifetime work interruptions and examines the relationship between work interruptions and earnings. Descriptive data showing the extent to which men and women have experienced work interruptions are presented, followed by an analysis of the impact of work interruptions on earnings. The study concludes that work interruptions explain only a small proportion of the earnings differential between men and women.

The 1979 ISDP was a panel survey of approximately 9,000 households that were visited at 3-month intervals over a period of a year and a half beginning in February 1979. The survey, part of the development stage of the new income survey called the Survey of Income and Program Participation (SIPP), was a joint effort of the U.S. Department of Health and Human Services and the U.S. Bureau of the Census. The third wave questionnaire contained a section on personal history and within that section were questions on lifetime work interruptions. The questions (reproduced ^{in Figure A} ~~in app. C~~) asked whether the person had ever been away from work for 6 months or longer for each of three reasons: (1) because he or she was not able to find work, (2) because he or she was taking care of home or family, and (3) because he or she was ill or disabled. Beginning and ending dates were recorded for each interruption. A maximum of four interruption periods were identified for each of the three possible reasons for interrupting.

A major reason for the interest in data on lifetime work experience is the desire to use such data in the analysis of male-female earnings differentials. The tenets of human capital research have traditionally stressed the importance of work experience patterns as a determinant of earnings. The descriptive data presented in the first part of this report confirm that the lifetime labor force attachment of women is weaker than that of men. Because of interruptions for familial reasons, women have a much higher overall rate of work interruptions than men and they spend a much higher proportion of their potential work years out of the labor force. Such findings have led at least some social scientists to posit that traditional familial responsibilities are one major reason why women earn less than men. This section will describe selected studies of the relationship between work interruptions and earnings and will present an analysis based on the 1979 ISDP data.

Previous Research

A major constraint in early efforts to examine the relationship between experience and earnings was the lack of data on lifetime work experience. More recently, however, a number of studies have been published which exploit the important data which has been made available from the National Longitudinal Surveys of Labor Market Experience (NLS) and the Michigan Panel Survey of Income Dynamics (PSID).

Suter and Miller (1973) were among the first to analyze the retrospective work history data from the NLS. They studied a cohort of women who were 30 to 44 years of age in 1967. Work experience was based on a question which asked about the total number of years in which the person

had worked at least 6 months. Suter and Miller concluded that there was a close association between earnings and length of work experience among this cohort of women.

Mincer and Polachek (1974) extended the analysis of the NLS retrospective data. They specified two reasons why discontinuous work history patterns might lead to lower earnings. First, interruptions in market work lead to lower levels of accumulated human capital. Second, interruptions cause a depreciation in existing human capital. That is, time spent away from market work has a cost beyond the effect of foregone experience. In their analysis, Mincer and Polachek found that the amount of time spent at home had a negative impact on earnings even when experience was also included in the earnings equation. They concluded from their analysis that a depreciation effect does, in fact, exist.

This finding was challenged by Sandell and Shapiro (1978) on the grounds that the NLS data used by Mincer and Polachek were subject to various coding errors. They replicated certain of the Mincer-Polachek research using a corrected NLS file and concluded that the original study had overestimated the depreciation effect.

Corcoran (1979) conducted an analysis of the effect of experience and interruptions on earnings using retrospective data from the PSID. One of the major advantages of the PSID data set was that the sample, in contrast to the NLS samples, was representative of the female population 18 to 64 years of age. Corcoran found very little evidence of a depreciation effect. There was no effect for White women and only a minor effect for Black women. Corcoran also argued that restricting the analysis group to women 30 to 44 years of age is likely to overestimate depreciation because

many women in this group have recently reentered the labor market and are likely to be affected by misinformation about job opportunities.

More recently, Mincer and Ofek (1982) used NLS data for 30- to 44-year-old married women to reaffirm the depreciation hypothesis. In an analysis of longitudinal (rather than retrospective) data from the NLS, they found that reentry wage rates were lower than wage rates at the time of labor force withdrawal. Furthermore, longer interruptions carried greater wage penalties. They also found, however, that wage rates tended to grow rapidly upon return to work. The observed amount of depreciation, they concluded, is dependent upon the length of the interruption and the length of time spent back in the labor force.

ISDP Data

The effect of work interruptions on earnings was examined by using the data described earlier to construct variables representing interruptions and experience. These variables were included in regressions which related hourly earnings to a set of explanatory variables. The universe for this part of the study consisted of all persons 21 to 64 years of age with wage and salary income during the quarter preceding the interview. Separate regressions were run for men and women, with the log of hourly earnings as the dependent variable.¹

The interruption and experience variables used in the regressions include the following:

- UNEMP = 1 if person had ever experienced an interruption due to an inability to find a job; 0 otherwise.
- DISAB = 1 if person had ever experienced an interruption due to illness or disability; 0 otherwise.

¹ Hourly earnings were calculated by dividing total earnings for the 3-month period by the total number of hours worked.

TIME-AWAY = Duration of all interruptions² as proportion of potential work years.³

EXPER = Number of potential work years minus duration of all interruptions.⁴

EXPERSQ = The square of EXPER

FT = 1 if the jobs the person has worked at have usually or always been full-time jobs; 0 otherwise.

The interruption variables were specified in the above form because it was hypothesized that earnings could be affected by the existence of an interruption as well as by the length of an interruption. Because interruptions due to unemployment or disability had a relatively small effect on the proportion of potential work years spent away from work, they were entered as zero-one dummy variables. Because interruptions for familial reasons had a very strong effect on the amount of time spent away from work, they were allowed to enter the equation through their effect on the TIME-AWAY variable. The general experience variable, EXPER, was entered in its own form as well as in its squared form, EXPERSQ. The inclusion of the squared form was intended to capture the non-linear effect of experience on earnings. (The returns to experience tend to flatten after some point.)

The education variables included in the regression were designed to take advantage of the ISDP personal history questions on highest

²A maximum of four interruption periods could be identified for each of three possible reasons for interrupting.

³Potential work years were defined as age minus years of school completed minus 6.

⁴The ISDP data on employer-specific or job-specific measures of work experience (e.g., tenure with most recent employer/at most recent job) were collected in the fifth wave of the survey and were not available for this study.

degree obtained, vocational training, and types of courses taken in high school. They included the following:

- EDUC1 = With an advanced degree
- EDUC2 = With a bachelors' degree
- EDUC3 = High school graduate (reference group)
- EDUC4 = Not a high school graduate, with a vocational training certificate
- EDUC5 = Not a high school graduate, no vocational training certificate
- COURSES = Number of selected academic courses completed in high school

Finally, a set of variables representing marital history were included:

- MARR1 = Married, no marital disruption (reference group)
- MARR2 = With a marital disruption (ever widowed, divorced or separated)
- MARR3 = Never married

The means for all variables are presented in table C and the regression results are shown in tables D and E. The unstandardized regression coefficients (table D) represent the earnings return to variables included in the equations. The standardized coefficients shown in table E are computed on values of dependent and independent variables which have been standardized so that each variable has a standard deviation equal to 1. This technique makes it possible to use the size of the coefficient as the basis for comparing the relative importance of each of the variables in a given equation. Results are shown for White women and men as well as for all women and men in order to facilitate comparisons with previous studies.

Results are also shown for men and women 30 years of age and over with no familial interruptions as an alternative method of examining the influence of work interruptions.

The large differences between the sexes in the degree of work attachment are highly visible in table C. Men had, on the average, about 19 years of work experience and had spent only about 2 percent of their potential work years away from work. Women, on the other hand, had 14 years of work experience and had spent about 20 percent of their potential work years away from work. There were small or insignificant differences between men and women in the mean values of the other experience and interruption variables and in the mean values of most of the education and marital history variables. Men, however, were more likely than women to have received advanced degrees and a larger proportion of women than men experienced. The average hourly earnings of all women was \$4.38, about 63 percent as high as the average hourly earnings of \$6.92 for all men.

The regression results confirm the importance of experience as a determinant of earnings. The general experience variables EXPER and EXPERSQ are highly significant for both men and women (table D) and are important relative to other variables in the determination of hourly earnings (table E). Attachment to full-time work also has a significant effect on earnings. The coefficients of the experience variables show that the returns to experience are greater for men than for women.

The interruption variables, in general, have a negative effect on earnings, but the effect is not particularly strong or consistent. The coefficient of TIME-AWAY is significant for both men and women in the equation for persons of all races, but is significant for women only in

the equations for White men and women. Interruptions due to illness or disability (DISAB) have a significant negative effect on earnings in five of the equations, but interruptions due to inability to find work have a significant negative effect in only two of the equations.

That an earnings equation contains both experience and interruption variables that are significant is evidence that a depreciation effect does exist. In the equation for men of all races, the experience variables EXPER, EXPERSQ, and FT are highly significant and the interruption variables UNEMP and DISAB are also significant. In the equations for women of all races, the experience variables and the interruption variable TIME-AWAY have highly significant effects on earnings. The conclusion is that a depreciation effect does exist and information about work interruptions will improve those models which attempt to explain earnings.

The coefficients of the education and marital history variables are in line with expectations, but two findings should be noted. First, the coefficient of EDUC4 for men is less negative than the coefficient of EDUC5. This finding suggests that a vocational training certificate has a positive effect on earnings. Second, the coefficient of COURSES is highly significant even though other measures of educational attainment are also present in the equation. That is, for the purpose of explaining earnings, it is important to know about the types of courses taken in high school even when we already have information about years of school completed and highest degree obtained.

Table C shows that the mean earnings of women are only about 62 percent of the earnings of men even when the group under study is comprised of persons 30 years of age and over with no familial interruptions. This differential exists even though women in this universe have approximately the same mean years of experience as men. Table D shows why the large differential exists

even when the mean values of experience are so close. Among the men in this group, the coefficient of EXPER is highly significant, but among women, the coefficient is not significant.

In general, the standardized regression coefficients (table E) reveal that the work interruption variables are less important than either the general experience or education variables as determinants of earnings. This holds true for both men and women. So, while the work interruption variables do show that a depreciation effect exists, general work experience and education are more critical determinants of earnings.

The earnings equations which have been developed for this report can be used to examine the extent to which differences in work history (experience and interruptions) are related to the earnings gap between men and women. That is, given the coefficients of their own equation, what would the earnings of women be if they had the same mean values as men for the variables measuring experience and interruptions. Table F shows that the effect of assigning to women the mean experience and interruption values of men is to reduce the earnings gap by only 12 percent.

Problems in retrospective measures of work experience and work interruptions

One of the goals of SIPP is to develop a data base that can be used to investigate the relationships among income, program participation, and personal history including work history. A certain amount of work history will be obtained as persons are followed during their time in the panel, but persons spend only 2 1/2 years in the panel. Some work history data may be obtained by matching survey records with Social Security earnings records, but matching records takes time and the amount of work history data that can be obtained

from Social Security records are limited. Until 1978, the Social Security record contained information on earnings during a quarter which were subject to the Social Security tax. Therefore, if a person's earnings met the Social Security tax limit in the first quarter of the year, no earnings data would appear for the remaining quarters. Since 1978, the record contains annual data on covered and noncovered earnings.

When the personal history supplement was designed for the third wave of the 1979 ISDP, the problem was to develop a set of questions which could be completed in 2 or 3 minutes and which would provide an indication of lifetime work attachment. The approach adopted was to attempt to identify periods lasting 6 months or longer when the person did not work. The ISDP work history questions are reproduced in Figure A.

There are obviously very great problems in trying to measure lifetime work experience in a brief set of questions. The data from these questions do seem to have a considerable amount of face validity, but it is certain that the data are also characterized by serious response problems. One way of identifying possible problem areas is to cross-classify current age by age at first reason-specific interruption. If there is no significant memory loss, then one would expect that the proportion of persons reporting that a first reason-specific interruption took place while they were in a particular age interval would be independent of their current age. Tables G and H show that memory loss is a significant factor in the reporting of first interruptions due to an inability to find work. Persons 21 to 29 were much more likely than older persons to report that such an interruption occurred before their 25th birthday. There is some evidence of memory loss in the reporting of first-time interruptions due to disability, but not to the same

degree as interruptions due to an inability to find work. There is no evidence of memory loss in the reporting of first-time interruptions of female interruptions for familial reasons. (The above conclusions are based on the assumption that the age groups had similar experiences.)

The ISDP results were taken into consideration when the time came to design the SIPP questions on work history (see Figure B). In an effort to reduce the problem of memory loss, respondents were asked to begin with the earliest 6 month interruption and work forward. The sequence also attempted to determine the total number of interruptions, then, for each period of interruption, determine the duration of and reason for the interruption. Because the SIPP sequence asks for the total number of interruptions and contains a "Don't Know" box for duration of interruption, we expect to be able to do a better job of imputing for item nonresponse.

Future work

We have finished the field collection operation for the third wave of SIPP, the wave containing the work history data. We are in the process of designing processing specifications so that a file can be prepared which contains no item nonresponse.

There are some differences between the work history data collected in ISDP and the SIPP work history data. First, the SIPP sample size is 20,000 households about twice the size of ISDP. Second, the SIPP data on beginning and ending dates of interruptions should be more complete than similar data from ISDP. Third, unlike the ISDP third wave file, the SIPP file will have data on job and occupation tenure. The SIPP file should be somewhat more useful than the ISDP file, and should allow users to expand the analysis by considering other variables (e.g. job and occupation tenure) and by considering the timing of work interruptions not just their existence and duration.

FIGURE A

ISDP Questions on Work Interruptions

<p>19b. Thinking about the jobs that . . . has worked at during . . .'s adult life, has . . . always worked full time, most often worked full time, most often worked part time or always worked part time?</p>	<p>3131 1 <input type="checkbox"/> Always full time - SKIP to 20a 2 <input type="checkbox"/> Most often full time 3 <input type="checkbox"/> Most often part time 4 <input type="checkbox"/> Always part time</p>																																
<p>20a. We would like to know about . . .'s experiences with unemployment. Have there been any times in . . .'s adult life when . . . was out of work for 6 months or longer because . . . could not find a job?</p>	<p>3134 1 <input type="checkbox"/> Yes 2 <input type="checkbox"/> No - SKIP to 21a</p>																																
<p>b. How often has this happened - just once, a few times, or many times?</p>	<p>3135 1 <input type="checkbox"/> Just once 2 <input type="checkbox"/> A few times 3 <input type="checkbox"/> Many times</p>																																
<p>c. In what years was . . . out of work for 6 months or longer because . . . could not find a job? (If more than 4 stretches, ask about most recent 4)</p>	<p>3136 * (1) <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> to <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> 3137 * (2) <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> to <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> 3138 * (3) <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> to <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> 3139 * (4) <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> to <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> 3140 7 <input type="checkbox"/> DK</p>	1	9			1	9			1	9			1	9			1	9			1	9			1	9			1	9		
1	9																																
1	9																																
1	9																																
1	9																																
1	9																																
1	9																																
1	9																																
1	9																																
<p>21a. Have there been times in . . .'s adult life when, instead of holding a job, . . . stayed home to take care of a family or home? (Include only periods lasting 6 months or longer)</p>	<p>3141 1 <input type="checkbox"/> Yes 2 <input type="checkbox"/> No - SKIP to 21c</p>																																
<p>b. In what years did . . . stay home to take care of a family or home? Was there any other time when . . . stayed home 6 months or longer to take care of a family or home?</p>	<p>3142 * (1) <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> to <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> 3143 * (2) <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> to <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> 3144 * (3) <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> to <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> 3145 * (4) <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> to <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> 3146 7 <input type="checkbox"/> DK</p>	1	9			1	9			1	9			1	9			1	9			1	9			1	9			1	9		
1	9																																
1	9																																
1	9																																
1	9																																
1	9																																
1	9																																
1	9																																
1	9																																
<p>c. Have there been times in . . .'s adult life when . . . did not work for 6 months or longer because . . . was sick or disabled?</p>	<p>3147 1 <input type="checkbox"/> Yes 2 <input type="checkbox"/> No - SKIP to 22a</p>																																
<p>d. In what years did . . . not work for 6 months or longer because . . . was sick or disabled? Was there any other time when . . . did not work for 6 months or longer because . . . was sick or disabled?</p>	<p>3148 * (1) <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> to <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> 3149 * (2) <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> to <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> 3150 * (3) <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> to <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> 3151 * (4) <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> to <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td>1</td><td>9</td><td></td><td></td></tr></table> 3152 7 <input type="checkbox"/> DK</p>	1	9			1	9			1	9			1	9			1	9			1	9			1	9			1	9		
1	9																																
1	9																																
1	9																																
1	9																																
1	9																																
1	9																																
1	9																																
1	9																																

Table A. Work Interruption History by Race, Spanish Origin, and Selected Characteristics: Males

Characteristic	Total (in thousands)	Percent with one or more interruptions lasting 6 months or more due to--				Mean percent of potential work years spent away from work for reasons surveyed	
		All reasons surveyed	Inability to find work	Family reasons	Illness or disability	Value	Standard error
Males 21 to 64 years who ever worked	55,828	25.7	17.3	1.5	10.6	3.3	0.1
RACE AND SPANISH ORIGIN ¹							
White	49,381	24.2	15.2	1.2	10.7	2.9	0.1
Black	5,627	40.2	35.0	3.9	10.7	6.9	0.3
Spanish origin	3,220	34.9	22.7	1.2	15.8	3.3	0.3
YEARS OF SCHOOL COMPLETED							
Less than 12	14,171	40.1	24.9	1.9	20.3	4.7	0.2
12 to 15	29,761	24.7	17.3	1.3	9.3	3.2	0.1
16 and over	11,896	11.0	7.9	1.6	2.4	1.7	0.2
AGE BY YEARS OF SCHOOL COMPLETED							
21 to 29 years	16,048	20.5	18.0	1.4	3.4	4.5	0.2
Less than 12	2,314	40.7	35.5	2.2	7.4	6.8	0.5
12 to 15	10,104	20.8	18.5	1.4	3.2	4.7	0.2
16 and over	3,630	6.9	5.5	.6	1.2	2.4	0.4
30 to 44 years	19,106	23.4	16.2	1.6	8.5	2.6	0.1
Less than 12	3,809	36.7	24.5	2.1	18.2	4.1	0.4
12 to 15	10,278	25.2	17.8	1.5	8.0	2.8	0.2
16 and over	5,019	9.6	6.5	1.4	2.0	.9	0.2
45 to 64 years	20,674	11.9	17.7	1.5	18.2	3.0	0.1
Less than 12	8,049	41.6	22.1	1.7	25.1	4.3	0.2
12 to 15	9,378	28.5	15.6	.8	17.1	2.1	0.2
16 and over	3,247	17.8	12.7	3.0	4.3	2.1	0.3
OCCUPATION GROUP OF USUAL JOB							
Professional, technical, or managerial	15,040	14.7	10.2	1.7	5.3	2.3	0.2
Sales or clerical	6,621	20.6	13.8	.9	7.3	2.3	0.2
Craftsmen	12,825	28.8	18.7	1.4	13.5	2.9	0.2
Operatives	10,254	32.5	20.8	.8	14.9	3.9	0.2
Laborers	5,832	37.9	27.6	2.2	13.6	5.6	0.3
Service	3,457	25.5	14.8	2.3	11.1	4.1	0.4

1/Persons of Spanish origin may be of any race.

Table B. Work Interruption History by Race, Spanish Origin, and Selected Characteristics: Females

Characteristic	Total (in thousands)	Percent with one or more interruptions lasting 6 months or more due to--				Mean percent of potential work years spent away from work for reasons surveyed	
		All reasons surveyed	Inability to find work	Family reasons	Illness or disability	Value	Standard error
Females 21 to 64 years who ever worked	57,258	71.9	14.2	64.1	9.2	30.9	0.2
RACE AND SPANISH ORIGIN¹							
White	49,812	73.0	12.4	66.8	8.3	32.7	0.2
Black	6,402	63.1	27.4	43.8	17.5	17.6	0.4
Spanish origin	3,014	75.0	23.6	62.4	12.9	27.6	0.6
YEARS OF SCHOOLING							
Less than 12	13,740	79.5	21.7	68.5	20.1	33.5	0.3
12 to 15	34,805	73.3	12.7	66.3	6.6	31.5	0.2
16 and over	8,713	54.3	8.6	48.6	2.6	24.2	0.4
AGE BY YEARS OF SCHOOL SCHOOL COMPLETED							
21 to 29 years	16,804	53.1	17.0	42.5	3.5	20.7	0.3
Less than 12	1,948	70.6	23.2	61.7	5.9	30.9	0.8
12 to 15	11,650	56.8	18.4	44.9	4.1	22.3	0.3
16 and over	3,206	29.1	8.3	22.2	.1	8.9	0.5
30 to 44 years	19,445	77.5	12.3	72.3	6.6	34.3	0.3
Less than 12	4,060	79.8	20.4	73.6	12.3	34.2	0.6
12 to 15	12,366	79.8	9.9	75.3	5.7	34.8	0.3
16 and over	3,018	65.1	11.4	58.5	2.9	32.2	0.7
45 to 64 years	21,011	81.7	13.8	73.8	16.1	35.8	0.3
Less than 12	7,733	81.5	22.0	67.6	27.8	33.7	0.4
12 to 15	10,789	83.7	9.8	79.0	10.3	37.7	0.4
16 and over	2,489	73.8	5.6	70.6	5.4	34.3	0.8
OCCUPATION GROUP OF USUAL JOB							
Professional, technical, or managerial	11,723	61.0	9.5	55.4	5.4	24.4	0.3
Sales or clerical	23,782	75.2	10.7	69.4	5.9	33.8	0.2
Operatives	8,447	74.8	22.2	62.9	14.5	29.4	0.4
Laborers	950	78.3	21.9	67.8	15.1	39.7	1.2
Service	10,543	74.2	19.7	63.4	16.5	32.1	0.4

¹/Persons of Spanish origin may be of any race.

Table C. Mean Values of Regression Variables

Variable	Males	Females	White males	White females	Males 30 and over with no familial interruptions	Females 30 and over with no familial interruptions
Interruption and experience variables:						
UNEMP135	.117	.121	.098	.137	.117
DISAB060	.049	.062	.040	.078	.093
TIME-AWAY017	.203	.015	.220	(X)	(X)
EXPER	19.256	14.334	19.337	13.868	25.223	24.644
EXPER SQ	535.421	331.227	537.208	312.682	755.661	745.445
FT969	.858	.968	.853	.990	.928
Education variables:						
EDUC1069	.044	.071	.045	.080	.076
EDUC2161	.128	.166	.135	.142	.107
EDUC4019	.015	.017	.009	.023	.020
EDUC5204	.157	.191	.145	.243	.274
COURSES	2.170	2.129	2.202	2.167	2.059	1.843
Marital history variables:						
MARR2214	.303	.204	.277	.255	.388
MARR3154	.172	.144	.159	.057	.203
Earnings variables:						
Hourly earnings	\$6.92	\$4.38	\$7.14	\$4.42	\$7.58	\$4.71
Log of hourly earnings .	1.934	1.478	1.966	1.487	2.026	1.550
Unweighted N	3,157	2,416	2,854	2,101	2,145	593

X Not applicable.

Table D. Coefficients of Regression of Log of Hourly Earnings on Specified Explanatory Variables

(Standard errors in parentheses)

Variable	Employed males	Employed females	Employed White males	Employed White females	Males 30 and over with no familial interruptions	Females 30 and over with no familial interruptions
UNEMP	-.039 (.018)	.002 (.018)	-.028 (.021)	.002 (.021)	-.078 (.018)	.014 (.041)
DISAB	-.125 (.023)	-.040 (.028)	-.144 (.025)	-.088 (.032)	-.143 (.025)	-.183 (.044)
TIME-AWAY	-.312 (.122)	-.128 (.025)	-.068 (.145)	-.155 (.028)	(X) -	(X) -
EXPER03515 (.00175)	.02278 (.00184)	.03791 (.00189)	.02495 (.00200)	.03382 (.00306)	.00937 (.00600)
EXPER SQ	-.00058 (.00005)	-.00042 (.00005)	-.00065 (.00005)	-.00046 (.00005)	-.00056 (.00005)	-.00014 (.00012)
FT216 (.032)	.112 (.016)	.254 (.035)	.099 (.018)	.363 (.064)	.372 (.048)
EDUC1336 (.023)	.358 (.028)	.338 (.023)	.322 (.030)	.327 (.028)	.301 (.053)
EDUC2179 (.016)	.218 (.018)	.181 (.018)	.209 (.021)	.231 (.021)	.260 (.046)
EDUC4	-.069 (.039)	-.146 (.048)	-.002 (.044)	-.120 (.067)	-.026 (.044)	-.415 (.092)
EDUC5	-.195 (.016)	-.190 (.018)	-.173 (.016)	-.179 (.018)	-.185 (.018)	-.244 (.035)
COURSES038 (.005)	.044 (.005)	.034 (.005)	.052 (.005)	.045 (.005)	.070 (.009)
MARR2	-.023 (.014)	.016 (.014)	-.038 (.014)	.038 (.014)	-.009 (.016)	-.035 (.030)
MARR3	-.192 (.016)	-.009 (.018)	-.141 (.018)	-.008 (.018)	-.279 (.030)	.029 (.035)
Constant	1.318	1.112	1.282	1.098	1.172	.993
R ²24	.18	.22	.19	.20	.28

- Represents zero.
X Not applicable.

Table E. Standardized Coefficients of Regression of Log of Hourly Earnings on Specified Explanatory Variables

Variables	Males	Females	White males	White females	Males 30 and over with no familial interruptions	Females 30 and over with no familial interruptions
UNEMP	-.026	.001	-.017	.002	-.050	.008
DISAB	-.056	-.018	-.067	-.037	-.073	-.098
TIME-AWAY	-.034	-.069	-.007	-.085	(X)	(X)
EXPER858	.546	.931	.582	.698	.203
EXPERT	-.653	-.419	-.735	-.438	-.628	.159
FT072	.084	.086	.075	.070	.178
EDUC1162	.158	.167	.142	.168	.147
EDUC2125	.156	.130	.152	.152	.148
EDUC4	-.018	-.038	-.001	-.024	-.007	-.108
EDUC5	-.149	-.148	-.131	-.134	-.150	-.201
COURSES131	.160	.118	.188	.154	.223
MARR2	-.018	.016	-.029	.036	-.008	-.031
MARR3	-.131	-.007	-.095	-.006	-.122	.021

Table F. Percent of the Male-Female Earnings Gap Explained By Experience, Interruption, and Education Variables

Variables	Employed persons all races	Employed persons White	Employed persons 30 years of age and over with no familial interruptions
Hourly earnings of men	\$6.92	\$7.14	\$7.58
Hourly earnings of women	\$4.38	\$4.42	\$4.71
Earnings gap	\$2.54	\$2.72	\$2.87
Hourly earnings of women adjusted for experience and interruption variables ¹	\$4.69	\$4.76	\$4.85
Difference between actual hourly earnings of women and hourly earnings adjusted for experience and interruption variables	\$.31	\$.34	\$.14
Percent of earnings gap explained by experience and interruption variables	12.2	12.5	4.9
Hourly earnings of women adjusted for experience, interruption, and education variables ²	4.75	4.80	5.64
Difference between actual hourly earnings of women and hourly earnings adjusted for experience, interruption, and education variables37	.38	.93
Percent of earnings gap explained by experience, interruption, and education variables	14.6	14.0	32.4

¹ Hourly earnings of women if women had the same mean values as men for experience and interruption variables. Experience and interruption variables include UNEMP, DISAB, TIME-AWAY, EXPER, EXPERSQ, and FT.

² Hourly earnings of women if women had the same mean values as men for experience, interruption, and education variables. Education variables include EDUC 1, EDUC 2, EDUC 4, EDUC 5, and COURSES.

Table G. Current Age by Age at Time of First Interruption: Males

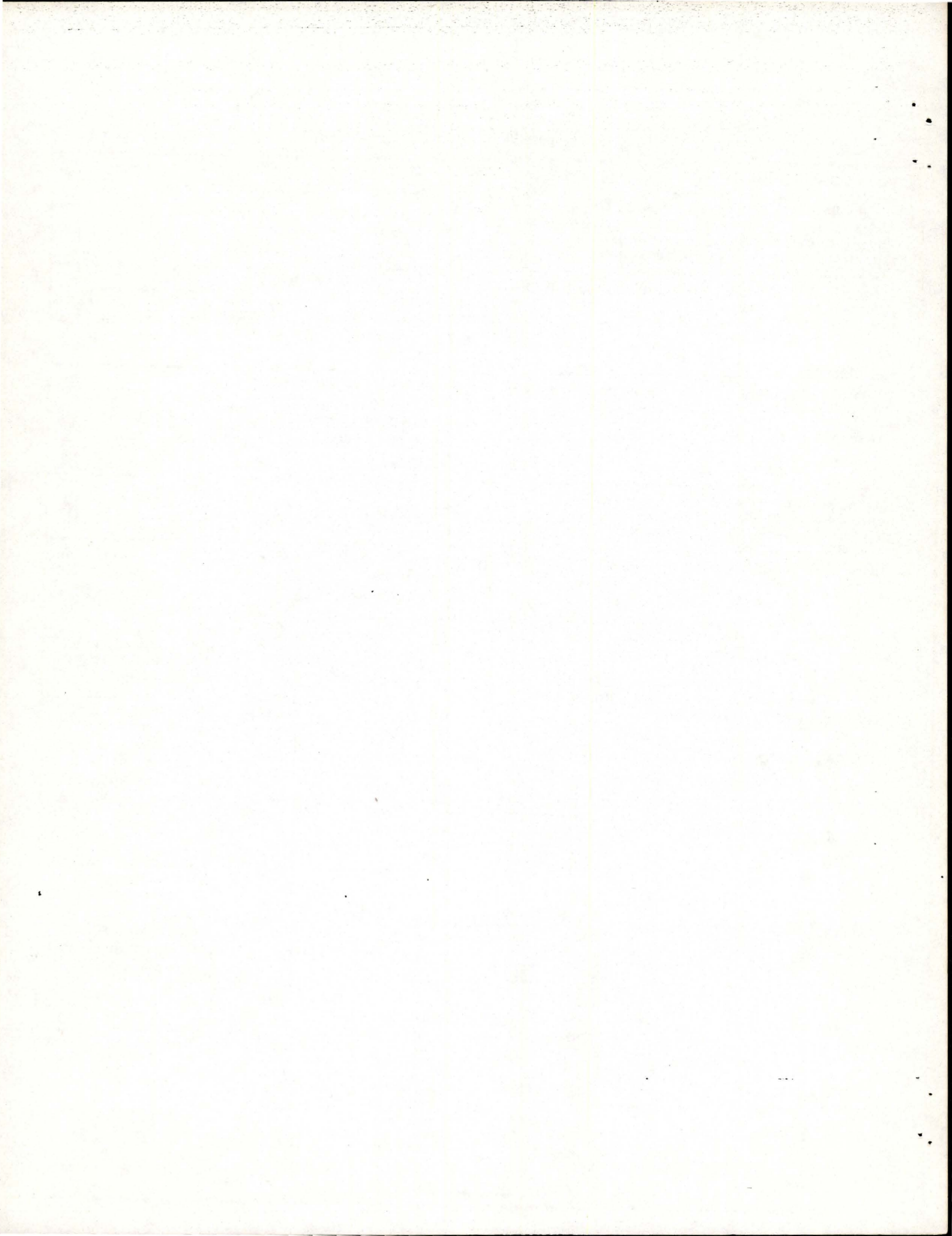
Reason for interruption and age at time of first interruption	Current age		
	21 to 29	30 to 44	45 to 64
All males	100.0	100.0	100.0
Age at time of first interruption due to inability to find work:			
Under 25	12.5	3.8	2.8
25 to 34	3.0	6.7	3.4
35 and over	-	2.4	7.7
NA	2.4	3.3	3.6
Age at time of first interruption due to disability:			
Under 25	2.6	2.0	1.7
25 to 34	0.4	4.3	3.0
35 and over	-	1.7	12.8
NA	0.4	0.4	0.7

Table H. Current Age by Age at Time of First Interruption: Females

Reason for interruption and age at time of first interruption	Current age		
	21 to 29	30 to 44	45 to 64
All males	100.0	100.0	100.0
Age at time of first interruption due to inability to find work:			
Under 25	12.6	2.9	1.6
25 to 34	2.0	5.5	2.0
35 and over	-	2.2	6.4
NA	2.5	1.8	3.7
Age at time of first interruption due to familial reasons:			
Under 25	32.5	38.5	34.2
25 to 34	8.6	25.9	21.8
35 and over	-	3.5	9.4
NA	1.4	4.3	8.5
Age at time of first interruption due to disability:			
Under 25	2.8	1.1	1.1
25 to 34	0.6	3.1	2.8
35 and over	-	1.6	9.9
NA	0.2	0.9	2.4

Figure B

<p>8a. People spend time out of the labor force for various reasons, such as taking care of a home or family, illness, going to school or other reasons. Since ... was 21 years of age, have there been any periods lasting 6 months or longer when ... did not work at a paid job or business?</p>	<p>8296 <input type="checkbox"/> Yes <input type="checkbox"/> No — SKIP to item 1, page 53</p>
<p>b. From the time ... was 21 years old, when was the first time that ... went 6 months or longer without working at a job or business?</p>	<p style="text-align: center;">FROM TO</p> <p>8298 <input style="width: 20px;" type="text" value="1"/> <input style="width: 20px;" type="text" value="9"/> <input style="width: 20px;" type="text"/> <input style="width: 20px;" type="text"/> 8300 <input style="width: 20px;" type="text" value="1"/> <input style="width: 20px;" type="text" value="9"/> <input style="width: 20px;" type="text"/> <input style="width: 20px;" type="text"/></p> <p>8302 x1 <input type="checkbox"/> DK</p>
<p>c. What was the reason ... did not work at a job or business during that time? <i>Mark (X) only one.</i></p>	<p>8304 <input type="checkbox"/> Took care of family or home <input type="checkbox"/> Own illness or disability <input type="checkbox"/> Could not find work <input type="checkbox"/> Going to school <input type="checkbox"/> Other</p>
<p>d. After this first time were there any other periods of 6 months or longer when ... did not work at a job or business?</p>	<p>8306 <input type="checkbox"/> Yes <input type="checkbox"/> No — SKIP to 1, page 53</p>
<p>e. How many other times did this happen?</p>	<p>8308 <input type="checkbox"/> One time <input type="checkbox"/> Two times <input type="checkbox"/> Three or more times</p>
<p><i>Ask 8f and 8g for each "Other" time: Maximum of three.</i></p> <p>f. When was the (second/third/fourth) time that ... went 6 months or longer without working at a job or business?</p> <p>SECOND TIME</p>	<p style="text-align: center;">FROM</p> <p>8310 <input style="width: 20px;" type="text" value="1"/> <input style="width: 20px;" type="text" value="9"/> <input style="width: 20px;" type="text"/> <input style="width: 20px;" type="text"/> TO</p> <p>8312 <input style="width: 20px;" type="text" value="1"/> <input style="width: 20px;" type="text" value="9"/> <input style="width: 20px;" type="text"/> <input style="width: 20px;" type="text"/></p> <p style="text-align: center;">OR</p> <p>8314 x1 <input type="checkbox"/> DK</p>
<p>THIRD TIME</p>	<p style="text-align: center;">FROM</p> <p>8318 <input style="width: 20px;" type="text" value="1"/> <input style="width: 20px;" type="text" value="9"/> <input style="width: 20px;" type="text"/> <input style="width: 20px;" type="text"/> TO</p> <p>8320 <input style="width: 20px;" type="text" value="1"/> <input style="width: 20px;" type="text" value="9"/> <input style="width: 20px;" type="text"/> <input style="width: 20px;" type="text"/></p> <p style="text-align: center;">OR</p> <p>8322 x1 <input type="checkbox"/> DK</p>
<p>FOURTH TIME</p>	<p style="text-align: center;">FROM</p> <p>8326 <input style="width: 20px;" type="text" value="1"/> <input style="width: 20px;" type="text" value="9"/> <input style="width: 20px;" type="text"/> <input style="width: 20px;" type="text"/> TO</p> <p>8328 <input style="width: 20px;" type="text" value="1"/> <input style="width: 20px;" type="text" value="9"/> <input style="width: 20px;" type="text"/> <input style="width: 20px;" type="text"/></p> <p style="text-align: center;">OR</p> <p>8330 x1 <input type="checkbox"/> DK</p>
<p>8g. What was the main reason ... did not work at a job or business during that time?</p>	
<p>8316 <input type="checkbox"/> Took care of family or home <input type="checkbox"/> Own illness or disability <input type="checkbox"/> Could not find work <input type="checkbox"/> Going to school <input type="checkbox"/> Other</p>	
<p>8324 <input type="checkbox"/> Took care of family or home <input type="checkbox"/> Own illness or disability <input type="checkbox"/> Could not find work <input type="checkbox"/> Going to school <input type="checkbox"/> Other</p>	
<p>8332 <input type="checkbox"/> Took care of family or home <input type="checkbox"/> Own illness or disability <input type="checkbox"/> Could not find work <input type="checkbox"/> Going to school <input type="checkbox"/> Other</p>	



PANEL SURVEYS AS A SOURCE OF MIGRATION DATA

Donald C. Dahmann

Population Division
U.S. Bureau of the Census
Washington, D.C. 20233

For presentation at the Annual Meeting
of the American Statistical Association,
Philadelphia, Pennsylvania, 13-16 August 1984.

INTRODUCTION

Panel surveys¹/ such as the Survey of Income and Program Participation now being conducted by the U.S. Bureau of the Census represent a natural source of data for the investigation of migration and other forms of geographical mobility. Information on the movement of individuals and households can be obtained directly through the process of maintaining the panel, making most supplementary questions for the specific purpose of obtaining migration data redundant and unnecessary. To address a wide range of geographical mobility research questions, it is important that information of sufficient geographic specificity be collected on the successive locales to which panel members relocate (e.g., address of place of residence, political jurisdiction, etc.), but then these same requirements exist for survey administration.

In at least one sense, we might expect that migration analysts would feel particularly comfortable with panel data, perhaps even more so than others who are now coming to use such data. One would expect this to be the case because the temporal dimension is an inherent aspect of both geographical mobility processes and panel surveys. Geographical mobility--whether residential mobility or migration--has always been framed in longitudinal terms, by asking the basic question: "Which actors (individuals, households, etc.) are in one state (previous residence, some portion of a settlement, region, etc.) at one time, and in another state at a subsequent time?" Individuals (households) identified as having changed states between two reference dates are movers; those in the same state at both beginning and end of the period are non-movers.

From this basic data-acquisition question, information is provided on the various forms of geographical mobility--residential, internal migration, international migration, etc.--and the characteristics of those who move.

Such information has been found invaluable for understanding changes in the nature of housing markets, labor markets, settlement systems, and the like.

A variety of forces bringing about major adjustments in the economic and social structure of the nation are serving to increase the demand for geographical mobility data. These forces include substantial shifts in the demographic composition of the nation's population (lower birth rate but increasing numbers of births, increasing numbers of elderly, decreasing family sizes, passage of the post World War II baby boom cohort into adulthood), nation-wide economic slowdown that is highly differentiated regionally (unemployment, slackening sales, etc.), and significant fluctuations in factors affecting supply-side components of housing markets (tenure conversions, building moratoriums, high interest rates). Each of these forces is spatially differentiated and thus producing geographically based impetus for population redistribution through migration. The nation is also becoming increasingly aware that as overall population growth due to natural increase continues to diminish, geographical mobility comes to play an increasing role in determining the growth and decline of individual locales, markets, and regions of the country.

In spite of significant increases in the demand for understanding geographical mobility processes (with accompanying increased data demands), rather little use has been made of panel surveys in geographical mobility research. Such a paradox can be understood only when it is realized that migration analysts regularly have available, and make use of, retrospective mobility data.^{2/} A multitude of surveys, as well as the decennial Censuses of Population and Housing, regularly produce geographical mobility data by inquiring of a respondent's previous place of residence, one year ago, five years ago, during adolescence, at time of birth, when a household moved into a

dwelling unit, etc.

For the most part, such retrospective data have served geographical mobility researchers admirably, although the general limitations--clustering of events at cardinal dates, memory loss, etc.--of this approach to obtaining survey data are well known. More important than data quality problems associated with retrospectively obtained information, however, is the fact that a full complement of retrospective information is rarely obtained. This makes it extremely difficult to ascertain the joint-incidence of migration with other events such as new household formation, renters purchasing a home of their own, a change in employment status, and similar major life events.

Even though panel data have previously been rather little utilized for geographical mobility research, they represent a data form of major potential that one can expect will be turned to as knowledge of both their availability and results of research utilizing them become more generally available. The remainder of this paper explores: (1) some major examples of previous uses of panel data, and (2) prospective uses with SIPP. The value of previous panel surveys for geographical mobility research is demonstrated with examples from two of the major social experiments conducted during the 1970s--Income Maintenance Experiment and Experimental Housing Allowance Program--and one of the major social science data collection efforts that was designed as a panel survey--Panel Study of Income Dynamics. Following this, the nature of the Survey of Income and Program Participation is briefly detailed, and that survey's potential for geographical mobility research addressed.

PREVIOUS USES OF PANEL SURVEYS FOR GEOGRAPHICAL MOBILITY RESEARCH

Panel surveys, the technique of using repeated interviews with a sample of individuals that remains constant, have been employed as a research strategy

to assist in understanding processes with an inherent capacity for change for at least half a century. Virtually all of the earliest uses of panel analysis were employed to investigate change in opinions, and particularly in preferences for political candidates (Sharma, 1975; Levenson, 1978; Rice, 1928; Lazarsfeld and Fiske, 1938; and Berelson, Lazarsfeld, and McPhee, 1954). From these beginnings, various forms of panel analysis, and more generally longitudinal analysis, have come to be utilized in a wide variety of social science research applications, and recently were singled out by the Office of Federal Statistical Policy and Standards as an approach to data collection and analysis to which "much attention should be devoted. . .in the development of [federal] statistical programs for the 1980's" (U.S. Department of Commerce, 1978; p. 321).

The real blossoming of panel surveys, both in terms of numbers of surveys and their size, occurred during the past two decades (and principally during the 1970s). They were employed as a means of monitoring and evaluating the effects of a variety of federally sponsored large-scale experiments and programs, and to enhance our understanding of the factors involved in a variety of social and economic processes, e.g., educational attainment, socialization, and labor force participation.

The principal federally sponsored experiments including panel analysis components were the several Income Maintenance Experiments, the various Experimental Housing Allowance programs, and the Urban Homesteading Demonstration. Uses of panel analysis in each of these programs for geographical mobility research are discussed in the next section.

Several large-scale panel surveys were also initiated during this period in response to the need for basic information on educational attainment, labor force participation, and social stratification processes (Borus, 1982). Major

surveys in this second group include the National Longitudinal Survey of Labor Market Practices, begun in 1966 and sponsored by the U.S. Department of Labor with distribution through Ohio State University's Center for Human Resource Research (popularly referred to as the Parnes data; Parnes, 1974; Center for Human Resources Research, 1974; Bielby, Hawley, and Bills, 1977; Parnes and Rich, 1980; Leigh, 1982; Daymont, 1983); the Panel Study of Income Dynamics, conducted by the University of Michigan's Institute for Social Research for the U.S. Department of Health and Human Services since 1968 (Duncan and Morgan, 1982); Project TALENT, conducted by the American Institutes for Research (Wise and Steel, 1980); Youth in Transition Project, conducted by the University of Michigan's Institute for Social Research (Bachman and O'Malley, 1980); National Longitudinal Survey of High School Seniors, conducted by Research Triangle Institute for the National Center for Education Statistics (Eckland and Alexander, 1980); the Wisconsin Longitudinal Study (Sewell and Hauser, 1980); and Explorations in Equality of Opportunity, initiated by the Educational Testing Service (Alexander and Eckland, 1980)3/.

Questions that these latter surveys were designed to address rarely included geographical mobility as a major topical area. Much attention in terms of questionnaire design or questions asked however, was not required as geographical mobility data flowed as a natural outcome of the follow-up of panel members. Thus, general information on geographical mobility derived of the panel design was sufficient for relating movement with labor force participation, educational attainment, social and occupational stratification processes, household changes with passage through the life course, shifts in housing consumption, etc.

The Experimental Housing Allowance Programs and the Income Maintenance Experiments, on the other hand, were both specifically interested in geograph-

ical mobility as an integral element of evaluating the effectiveness of the programs. Both programs, for instance, were concerned with the effect upon, and role of, residential mobility in relation to patterns of housing consumption, and the Income Maintenance Experiments were further concerned with the role of migration as a response to income guarantees among lower-income households. The fact that information on the various forms of geographical mobility flowed directly from the panel design of these surveys without specific geographical mobility questionnaire items demonstrates the capacity of panel surveys to contribute toward our understanding the role of geographical mobility in a variety of social and economic circumstances.

Experimental Housing Allowance Program

The Experimental Housing Allowance Program, initiated by the Housing and Community Development Act of 1970, was undertaken to establish empirical evidence of the effects of housing allowances, and of the transfer of (small amounts of) unrestricted funds to lower-income households on housing consumption. The following questions highlight some of the experiment's major research goals. Would households spend the money on housing? Would the money be used to improve conditions of their current dwelling? Would households move to other neighborhoods? What would be the local housing market impact of such an infusion of funds? To answer such questions, three programs--Housing Allowance Supply Experiment, Housing Allowance Demand Experiment, and Administrative Agency Experiment--eventually enrolled more than 30,000 households at twelve sites throughout the country at a cost in excess of \$160 million (Friedman and Weinberg, 1982; 1983; Bradbury and Downs, 1981; Struyk and Bednick, 1981).

Evaluation of these individual programs produced several longitudinal

analyses, but only one that utilized individual households as its unit of observation. Other longitudinal analyses took housing units and neighborhoods as their units of observation (Hillestad and McDonald, 1983), or obtained retrospective geographical mobility data for individuals (McCarthy, 1983). Individuals in both control and test groups of the Housing Allowance Demand Experiment were traced for three years to observe, among other things, actual patterns of residential mobility and what changes, in terms of housing consumption and residential dispersion (and therefore desegregation), may have resulted from such moves (Rossi, 1981; Hamilton, 1983).

Similar panel analyses were utilized in evaluating the effects of one other important housing experiment initiated at about the same time, the U.S. Department of Housing and Urban Development's Urban Homesteading Demonstration. This effort, begun as a demonstration and later established as a regular program, transferred HUD-owned properties to local control in 23 cities (U.S. Department of Housing and Urban Development, 1977). Its research agenda included inquiry into the effects of residential mobility on patterns of local housing consumption, this time with specific reference to the displacement of low-income households from housing they could no longer afford as the result of the HUD-owned properties being returned to the market (Schnare, 1979).

Income Maintenance Experiments

Another of the recent massive social experiments that included panel analysis as a research component is the series of income maintenance experiments, the first large scale social experiments to be conducted in the United States (begun in 1967)—New Jersey Income Maintenance Experiment (Kershaw and Fair, 1976; Watts and Rees, 1977; Pechman and Timpane, 1975), Rural Income

Maintenance Experiment (Bawden and Harrar, 1976; Palmer and Pechman, 1978), Gary Income Maintenance Experiment (Journal of Human Resources, 1979), and the Seattle and Denver Income Maintenance Experiment (Journal of Human Resources, 1980). Each of these four programs addressed various aspects of one basic issue: how much would a nationwide guaranteed income cost, and to what extent would families reduce their labor force participation (and therefore earnings) in response to such payments?

As with the massive housing allowance experiments, research agendas of the income maintenance experiments did not include geographical mobility as a specific focus. Nonetheless, the panel designs employed in the evaluations, which traced families (households) over a three-year period, produced their own mobility data. Analyses were undertaken of both of the two major forms of geographical mobility: (1) migration, specifically rates of movement from the experimental site to other labor markets and, (2) residential mobility, change of one's dwelling to consume a different bundle of housing services (quality of dwelling, neighborhood services, etc.). Data to investigate both of these extremely important outcomes of the decision to move were readily derivable from the panel design of the surveys used to evaluate the experiments.

Findings with regard to migration (spatial adjustments in labor force participation) may be summarized as follows: (1) migration out of the experimental site's labor market was significantly increased for married white males and females but not for married black males and females and, (2) outmigration was to locales with generally lower wage rates and with better living environments. Work hours in the new locations were generally less than previously, suggesting either that persons worked fewer hours in their new locations because of their additional income or that their search for a "satisfactory" job in the new locale took some time (Keeley, 1980).

With regard to residential mobility, it was discovered that (1) households moving to a new address generally improved their housing situation (Wooldridge, 1977; Kaluzny, 1979), and (2) the effects of income assistance as a means of enabling renter households to move into a home of their own were mixed (Wooldridge, 1977; Poirier, 1977).

Panel Study of Income Dynamics

A more archetypical panel study, at least in traditional terms, is the Panel Study of Income Dynamics (PSID) conducted by the Institute for Social Research of the University of Michigan for the U.S. Department of Health and Human Services. Initiated during the same period--Great Society Era of the 1960s--as the Income Maintenance Experiments, this panel survey is now in its 17th year of collecting annual information from a representative national sample of about 6,000 families and 15,000 individuals (Morgan and Smith, 1969). The Survey has produced a massive body of data, a massive array of findings (Duncan and Morgan, 1982; Morgan, 1974; Duncan and Morgan, 1975-1980; and Hill, Hill, and Morgan, 1981), and even outlived its original sponsoring federal agency, the Office of Economic Opportunity.

Panel Study of Income Dynamics data have also been utilized to address questions in both of the two major realms of geographical mobility research--migration and residential mobility. It has also been utilized to consider more basic geographic mobility research questions--timing of moves through the life course, relationships between desires or expectations to move and actual movement, and other similar questions that are intrinsic to the geographical mobility processes.

The rich set of personal attribute and attitudinal variables in the Michigan panel has enabled residential mobility research to be framed in behavioral

terms, whereby households are seen as possessing specific desires and preferences with respect to moving. Structural elements of the participation system--income levels and purchasing power, housing costs, forced relocation, etc.--in this frame of reference then serve to enable or hinder actual patterns of mobility, and therefore preference fulfillment (Roistacher, 1974; 1975; Goodman, 1974; Duncan and Newman, 1975; 1976; Newman and Duncan, 1979; and Newman and Owen, 1982).

Use of these panel data have also enabled researchers to examine the characteristics of migrants in various interregional migration systems during the 1970s (Kim, 1980), patterns and consequences of repeat moves (Newman and Ponza, 1981), and to initiate structuring of general models of mobility (Morgan, 1977). Migration as an act resulting in the readjustments of labor markets has also been explored both in terms of causes, such as the effects of unemployment on movement (DaVanzo, 1978), and of consequences for individuals, in terms of income and occupational change (Harris, 1981).

THE SURVEY OF INCOME AND PROGRAM PARTICIPATION AS A SOURCE OF MIGRATION DATA

The Survey of Income and Program Participation (SIPP), a general purpose, large scale (25,000 household), national representative sample survey, has been undertaken by the U.S. Bureau of the Census primarily to provide: (1) improved data on the economic situation of individuals and households and (2) information on federal and state income transfer and social program participation. Individuals are interviewed every four months for the life of a panel. In the case of the first (1984) panel, this will result in a total of 9 waves of interviews for three-quarters of the panel and 8 waves for the remaining quarter. Initial interviews for the 1984 panel were conducted in October 1983 (with a reference period of July-September 1983); the final wave

of interviews for the this panel will be conducted in May 1986 (with a reference period of January-April 1986). Current plans for a second (1985) panel calls for a somewhat smaller sample size (20,000 households) and eight waves of interviews, which will begin in January 1985.

The earlier review of geographical mobility research topics explored with data from panel surveys presages the types of research we might expect from SIPP. In terms of duration of the panel, SIPP data will be much like those that were derived of the several large scale experiments--both cover a period not exceeding three years. In terms of geographical mobility research therefore, data from both sources may be used to explore change over only a relatively short period of time. Several SIPP characteristics make it a close relative of the Panel Study of Income Dynamics as well: (1) its sample is national (though larger) rather than being limited to selected settlements, and (2) more waves of interviews are being conducted, which will provide better data for establishing joint-incidence of movement with other--life course, employment, etc.--events.

The geographical mobility research derived of the panel surveys discussed earlier suggest that attention must be given to two specific aspects of such surveys: first, the periodicity of waves and overall duration of the panel; and second, the substantive nature of data collected during each wave. SIPP is unique in the short span of time between waves--four months. This design characteristic makes it particularly valuable as a means of matching residential shifts and migration with other life events such as marriage, divorce, expansion and contraction of household size in general, loss of job, change in job, and the like. No previous national survey has provided such a fine temporal scale for establishing the joint-incidence of geographical movement with important employment and life events.

The fact that each panel collects data for 2 1/2 years presents both advantages and disadvantages. As SIPP panels will not be maintained for years unto decades, as have the several major panel surveys focusing on changes in labor force participation, educational attainment, and social mobility through the life course, analysis of the role, consequences, and duration of effects of geographical mobility through major stages of the life course is not feasible. Nonetheless, 2 1/2 years (plus the fact that a large number of waves will be conducted) is quite sufficient to establish both immediate and some intermediate-term effects associated with geographical mobility on topics of concern such as the spatial restructuring of labor markets. The duration of SIPP panels also provides a reasonable amount of time to relate the expectations of individuals regarding mobility to actual patterns of movement.

Once we have accustomed ourselves to the fact that the act of tracking those who move in a panel survey provides migration data, then what must be considered in addressing geographical mobility questions is the basic substance of the questionnaire administered prior to and following the move. In the case of SIPP we are provided with a wealth of relevant migration-related information: labor force participation and employment, industry and occupation, work history, education, health conditions and disability, household composition, and, of course, income. As the same questionnaires are administered at the same times to nonmovers, the opportunity exists for comparing the situations of movers and nonmovers directly.

In consideration of these several points, what should we be looking to SIPP for in terms of geographical mobility research? First, I think that we can expect better data.^{4/} For decades the Current Population Survey (CPS) has served as our national metric establishing levels of movement among the various components of the nation's settlement system, among regions, and among

subpopulations of the nation's peoples. All CPS geographical mobility data are collected retrospectively--sometimes asking respondents to refer to an event that occurred one year ago, sometimes five years ago. These data, like all retrospective data, are subject to biases introduced by the distorting effects of memory loss, dissonance reduction (rationalization), and the like. How does SIPP data compare with CPS data? Will SIPP's multiple waves of data collection enable us to specify the overall effects of repeat movers on mobility statistics in a way that cross-sectional data do not? What will differences between the two survey's geographical mobility data tell us?

The fact that information on movement (and non-movement) and a wide array of life events will be collected almost as they occur (specific to within four months) is one of SIPP's best features from the perspective of geographical mobility research. Our capacity to specify the relationships between movement and such events as the loss of a job, termination of the receipt of unemployment benefits, marriage, divorce, etc. has never been better.

A set of supplemental migration questions, which should be administered to all individuals for at least one (preferably early) wave of interviewing, should also be considered. First, respondents should be asked a set of mobility preference questions, to relate desires and expectations of movement with patterns of actual mobility events as revealed by the survey. Secondly, retrospective questions on one's general residential history should be asked so that subsequent moves may be related to previous patterns of mobility and locations.

One further aspect of SIPP's design that should not be overlooked when thinking about geographical mobility research is its capacity to provide information on the locales of origin and destination of movers (and nonmovers as well). The ability of SIPP to provide information on conditions in both

the labor markets that migrants leave, and those to which they move, is of particular concern when wishing to understand the spatial differentiation of labor markets and to ascertain the causes of subnational (regional) patterns of employment growth and decline. What, for instance, are the relationships between sending and receiving markets in terms of unemployment rates, wage levels, etc.? Are these structural situations consequential in terms of employment? Are different mechanisms operating for blue collar and white collar migrants that such differences articulate? These are some of the questions that should guide attempts to maximize the utilization of SIPP data for geographical mobility research.

CONCLUSION

The Survey of Income and Program Participation enables us to explore new questions concerning both of the major forms of geographical mobility—residential mobility and migration. With particularly good income and public program participation data, good specification of the timing of movement with significant life events, and (potentially) good market characteristics data, SIPP is ideally suited to address a multitude of housing consumption questions. With good information on participation in federal and state-sponsored programs, exceptionally good income data, and (potentially) good information on the characteristics of labor markets, SIPP promises to be an incomparable research tool for questions that have heretofore simply gone unasked regarding migration, and particularly as it relates to readjustments of the spatial dimension of labor markets.

We must also be prepared to take advantage of the serendipitous benefits of timing. In this regard, the availability of SIPP data and recent advances in analytical techniques provide us with opportunities that were nonexistent

even a decade ago. With regard to analytical techniques I am thinking particularly of those developed during the 1970s for analyzing categorical data (Goodman, 1978; Bishop, Fienberg, and Holland, 1975; Markus, 1979), and their specific application to the analysis of change in mobility and panel data (Hauser, 1979; Goodman, 1973; Duncan, 1979; 1981; Fienberg, 1980). The richness of SIPP data provide a wonderful opportunity to fully utilize the analytical advances brought about by these techniques to answer a myriad of geographical mobility questions.

The Great Society programs of the 1960s pushed social scientists as never before to ask questions about American society and its economy. In response to these demands, new and better data were collected, new analytical techniques were developed, and new research agendas established. Much was learned from these efforts about the causes, the roles performed by, and effects of geographical mobility on the nation's economic and social structure. SIPP represents a logical outcome of advances in social science data collection that began in the 1960s and an important new opportunity for geographical mobility research. I invite your comments on ways that we at the Census Bureau can enhance this new survey's utility for answering your geographical mobility questions.

Notes

1/ - Panel data are distinguished from other forms of data (e.g., cross-sectional) by two characteristics: (1) the unit of observation--individuals, households, dwelling units, neighborhoods, nations, etc.--remains constant over time, and (2) multiple observations (waves of interviews) are made on each unit. Panel data represent one form of longitudinal data, which more generally, meet only the first of the two defining criteria of panel data (i.e., the unit of observation remains constant over time).

2/ - Retrospective data is information about a respondent's status, action, or state at a previous point in time. It is collected by asking respondents to recall information about past events. Typical examples include place of birth, residence five years ago, marital history, dollar expenditures on household items over the past three months, etc. Evidence of the confusion that exists among these various forms of data (including longitudinal, cross-sectional, and retrospective data) and their role in migration research is presented in Shryock (1965).

3/ - Brief mention should be made of two other longitudinal (panel) surveys conducted by the Census Bureau. The Income Survey Development Program (ISDP), which served as the research and development phase of the Survey of Income and Program Participation (SIPP), is the first. The 1979 ISDP panel consisted of 9,300 households, two-thirds of which were interviewed six times and one-third five times (David, 1983; U.S. Bureau of the Census, 1982). The second, the Annual Housing Survey (AHS) retains individual dwellings as its units of observation rather than individual persons as in the ISDP and SIPP. Data from

its first panel (consisting of about 70,000 dwelling units) are publicly available for the years 1974 through 1981 (Abt Associates, 1983; U.S. Bureau of the Census, 1979). Two forms of geographical mobility data are available from the AHS files: (1) data on households that moved during the year prior to interview, which includes a rather full array of retrospective data on the household's previous housing situation, and (2) data on the succession of households residing in individual dwelling units (which remain the unit of observation). The first form of data are annually available as residential mobility tabulations (U.S. Bureau of the Census, 1983) and as computer data files; the latter data have been utilized to examine the racial succession of households in metropolitan housing during the 1970s (Dahmann, 1983) and are also available as computer data files (Abt Associates, 1983).

4/ - This conclusion is based in part on comparisons of Current Population Survey and Panel Study of Income Dynamics data for migration research (Bilsborrow and Akin, 1982).

References

- Abt Associates. 1983. Annual Housing Survey National Longitudinal File [Codebook]. Cambridge, Mass.: Abt Associates.
- Alexander, Karl L. and Bruce K. Eckland. 1980. "The 'Explorations in Equality of Educational Opportunity' Sample of 1955 High School Sophomores." In Alan C. Kerckhoff. ed. Research in Sociology of Education and Socialization. Vol. 1. Greenwich: JAI Press. Pp. 31-58.
- Bachman, Jerald G. and Patrick M. O'Malley. 1980. "The Youth in Transition Series: A Study of Change and Stability in Young Men." In Alan C. Kerckhoff. ed. Research in Sociology of Education and Socialization. Vol. 1. Greenwich: JAI Press. Pp. 127-160.
- Berelson, Bernard, Paul F. Lazarsfeld, and William N. McPhee. 1954. Voting: A Study of Opinion Formation in a Presidential Campaign. Chicago: University of Chicago Press.
- Bilsborrow, Richard E. and John S. Akin. 1982. "Data Availability versus Data Needs for Analyzing the Determinants and Consequences of Internal Migration." Review of Public Data 10 (December): 261-284.
- Bishop, Yvonne M. M., Stephen E. Feinberg, and Paul Holland. 1975. Discrete Multivariate Analysis: Theory and Practice. Cambridge: MIT Press.
- Bawden, Lee and William Harrar. eds. 1976. Rural Income Maintenance Experiment: Final Report. Madison: Institute for Research on Poverty, University of Wisconsin-Madison.
- Bielby, William T., Clifford B. Hawley, and David Bills. 1977. "Research Uses of the National Longitudinal Surveys." Special Report No. 18. Madison: Institute for Research on Poverty, University of Wisconsin.

- Borus, Michael E. 1982. "An Inventory of Longitudinal Data Sets of Interest to Economists." Review of Public Data Use 10 (May): 113-126.
- Bradbury, Katherine and Anthony Downs. eds. 1981. Do Housing Allowances Work? Washington, D.C.: Brookings Institution.
- Center for Human Resources Research. 1974. National Longitudinal Surveys Handbook. Columbus: Center for Human Resources Research, Ohio State University.
- Dahmann, Donald C. 1983. "Racial Differences in Housing Consumption During the 1970s: Insights from a Components of Inventory Change Analysis." Urban Geography 4: 203-222.
- DaVanzo, Julie. 1978. "Does Unemployment Affect Migration?--Evidence from Microdata." Review of Economics and Statistics 60: 504-514.
- David, Martin. 1983. Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (ISDP). New York: Social Science Research Council.
- Daymont, Thomas and Paul Andrisani. 1983. "The Research Uses of the National Longitudinal Surveys: An Update." Review of Public Data Use 11: 203-310.
- Duncan, Greg J. and James N. Morgan. 1975-1980. Five Thousand American Families--Patterns of Economic Progress. Volumes 3 to 8. Ann Arbor: Institute for Social Research, University of Michigan.
- Duncan, Greg J. and James N. Morgan. 1980. "The Incidence and Some Consequences of Major Life Events." In Greg J. Duncan and James N. Morgan. eds. Five Thousand American Families--Patterns of Economic Progress. Ann Arbor: Institute for Social Research, University of Michigan. Pp. 183-240.
- Duncan, Greg J. and James N. Morgan. 1982. "Longitudinal Lessons from the

- Panel Study of Income Dynamics." Review of Public Data Use 10: 179-184.
- Duncan, Greg J. and Sandra J. Newman. 1975. "People as Planners: The Fulfillment of Residential Mobility Expectations." In Greg J. Duncan and James N. Morgan. eds. Five Thousand American Families--Patterns of Economic Progress. Volume 3. Ann Arbor: Institute for Social Research, University of Michigan. Pp. 279-318.
- Duncan, Greg J. and Sandra J. Newman. 1976. "Expected and Actual Residential Mobility." Journal of the American Institute of Planners 42: 174-186.
- Duncan, Otis Dudley. 1979. "Testing Key Hypotheses in Panel Analysis." In Karl F. Schussler. ed. Sociological Methodology 1980. San Francisco: Jossey Bass. Pp. 279-289.
- _____. 1981. "Two Faces of Panel Analysis: Parallels With Comparative Cross-Sectional Analysis and Time-Lagged Association." In Samuel Linhardt. ed. Sociological Methodology 1981. San Francisco: Jossey Bass. Pp. 281-318.
- Eckland, Bruce K. and Karl L. Alexander. 1980. "The National Longitudinal Study of the High School Senior Class of 1972." In Alan C. Kerckhoff. ed. Research in Sociology of Education and Socialization. Vol. 1. Greenwich: JAI Press. Pp. 189-222.
- Ferber, Robert and Werner Z. Hirsch. 1978. "Social Experimentation and Economic Policy: A Survey." Journal of Economic Literature 16 (December 1978): 1379-1414.
- Fienberg, Stephen E. 1980. "The Measurement of Crime Victimization: Prospects for Panel Analysis of a Panel Survey." The Statistician 29: 313-350.
- Friedman, Joseph and Daniel H. Weinberg. 1982. The Economics of Housing Vouchers. New York: Academic Press.

- Friedman, Joseph and Daniel H. Weinberg. eds. 1983. The Great Housing Experiment. Urban Affairs Annual Review, Volume 24. Beverly Hills: Sage Publications.
- Goodman, John L. 1974. "Local Residential Mobility and Family Housing Adjustments." In James N. Morgan. ed. Five Thousand American Families--Patterns of Economic Progress. Volume 2. Ann Arbor: Institute for Social Research, University of Michigan. Pp. 79-105.
- Goodman, Leo A. 1973. "Causal Analysis of Data from Panel Studies and Other Kinds of Surveys." American Journal of Sociology 78: 1135-1191.
- _____. 1978. Analyzing Qualitative/Categorical Data. Cambridge, Mass.: Abt Books.
- Hamilton, William L. 1983. "Economic and Racial/Ethnic Concentration." In Joseph Friedman and Daniel H. Weinberg. eds. The Great Housing Experiment. Urban Affairs Annual Review, Vol. 24. Beverly Hills: Sage Publications.
- Harris, Richard J. 1981. "Rewards of Migration for Income Change and Income Attainment: 1968-1973." Social Science Quarterly 62: 275-293.
- Hauser, Robert M. 1979. "Some Exploratory Methods for Modeling Mobility Tables and Other Cross-Classified Data." In Karl F. Schussler. ed. Sociological Methodology 1980. San Francisco: Jossey Bass. Pp. 413-458.
- Hill, Martha S., Daniel H. Hill, and James N. Morgan. 1981. Five Thousand American Families--Patterns of Economic Progress. Volume 9. Ann Arbor: Institute for Social Research, University of Michigan.
- Hillstad, Carol E. and James L. McDowell. 1983. "Neighborhood Change." In Joseph Friedman and Daniel H. Weinberg. eds. Urban Affairs Annual Review, Vol. 25. Beverly Hills: Sage Publications.
- Journal of Human Resources. 1979. Issue devoted to "The Gary Income

- Maintenance Experiment." 14: 431-506.
- _____. 1980. Issue devoted to "The Seattle and Denver Income Maintenance Experiments." 15: 463-722.
- Kaluzny, Richard L. 1979. "Changes in the Consumption of Housing Services: The Gary Experiment." Journal of Human Resources 14: 496-506.
- Keeley, Michael C. 1980. "The Effect of a Negative Income Tax on Migration." Journal of Human Resources 15: 695-706.
- Kershaw, David and Jerilyn Fair. eds. 1976. The New Jersey Income-Maintenance Experiment. Vol. 1. New York: Academic Press.
- Kim, Joachul. 1980. "Characteristics of Migrants Within the Framework of Current Migration Direction in the United States: Some Evidence from Micro-Data Analysis." Policy Sciences 12: 355-378.
- Lazarsfeld, Paul F. and Marjorie Fiske. 1938. "The 'Panel' as a New Tool for Measuring Opinion." Public Opinion Quarterly 2: 596-612.
- Leigh, Duane E. 1982. "The National Longitudinal Surveys: A Selective Survey of Recent Evidence." Review of Public Data Use 10: 185-201.
- Levenson, Bernard. 1978. "Panel Studies." In William H. Kruskal and Judith M. Tanur. eds. International Encyclopedia of Statistics. New York: Free Press. Pp. 683-691.
- Markus, Gregory B. 1979. Analyzing Panel Data. Beverly Hills: Sage Publications.
- McCarthy, Kevin F. 1983. "Housing Search and Residential Mobility." In Joseph Friedman and Daniel H. Weinberg. eds. The Great Housing Experiment. Urban Affairs Annual Review, Vol. 24. Beverly Hills: Sage Publications.
- Morgan, James N. ed. 1974. Five Thousand American Families--Patterns of Economic Progress. Vols. 1 and 2. Ann Arbor: Institute for Social

Research, University of Michigan.

- Morgan, James N. 1977. "Some Preliminary Investigations for a Model of Mobility." In Greg J. Duncan and James N. Morgan. eds. Five Thousand American Families--Patterns of Economic Progress. Volume 5. Ann Arbor: Institute for Social Research, University of Michigan. Pp. 349-368.
- Morgan, James N. and James D. Smith. 1969. A Panel Study of Income Dynamics: Study Design, Procedures, and Forms--1968 Interviewing Year (Wave I) Vol. 1. Ann Arbor: Institute of Social Research, University of Michigan.
- Nelson, Dawn, David McMillan, and Daniel Kasprzyk. 1984. An Overview of the Survey of Income and Program Participation. SIPP Working Paper Series No. 8401. Washington, D.C.: U.S. Bureau of the Census.
- Newman, Sandra J. and Greg J. Duncan. 1979. "Residential Problems, Dissatisfaction, and Mobility." Journal of American Planning Association 45: 154-166.
- _____ and Michael S. Owen. 1982. Residential Displacement in the U.S., 1970-1977. Ann Arbor: Institute for Social Research, University of Michigan.
- _____ and Michael Ponza. 1981. "The Characteristics of Housing Demand in the 1970s: A Research Note." In Martha S. Hill, Daniel H. Hill, and James N. Morgan. ed. Five Thousand American Families--Patterns of Economic Progress. Ann Arbor: Institute for Social Research, University of Michigan.
- Palmer, John L. and Joseph A Pechman. eds. 1978. Welfare in Rural Areas: The North Carolina-Iowa Income Maintenance Experiment. Washington, D.C.: Brookings Institution.
- Parnes, Herbert S. 1974. "The National Longitudinal Surveys: New Vistas for Labor Market Research." American Economic Review 65: 244-249.

- _____ and Malcolm C. Rich. 1980. "Perspectives on Educational Attainment from National Longitudinal Surveys of Labor Market Behavior." In Alan C. Kerckhoff. ed. Research in Sociology of Education and Socialization Vol. 1. Greenwich: JAI Press. Pp. 161-188.
- Pechman, Joseph A. and P. Michael Timpane. eds. 1975. Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment. Washington, D.C.: Brookings institution.
- Poirer, Dale J. 1977. "The Determinants of Home Buying." In Harold W. Watts and Albert Rees. eds. The New Jersey Income-Maintenance Experiment. Vol. 3. New York: Academic Press. Pp. 73-91.
- Rice, Stuart A. 1928. Quantitative Methods in Politics. New York: Knopf.
- Roistacher, Elizabeth. 1974. "Residential Mobility." In James N. Morgan. ed. Five Thousand American Families--Patterns of Economic Progress. Vol. 2. Ann Arbor: Institute for Social Research, University of Michigan.
- Roistacher, Elizabeth. 1975. "Residential Mobility: Planners, Movers and Multiple Movers." In Greg J. Duncan and James N. Morgan. eds. Five Thousand American Families--Pattterns of Economic Progress. Vol. 3. Ann Arbor: Institute for Social Research, University of Michigan.
- Rossi, Peter H. 1981. "Residential Mobility." In Katherine Bradbury and Anthony Downs. eds. Do Housing Allowances Work? Washington, D.C.: Brookings Institution. Pp. 147-183.
- Rossi, Peter H. and Katherine C. Lyall. 1976. Reforming Public Welfare: A Critique of the Negative Income Tax Experiment. New York: Russell Sage Foundation.
- Schnare, Ann B. 1979. Household Mobility in Urban Homesteading Neighborhoods: Implications for Displacement. Washington, D.C.: U.S. Government Printing Office.

- Sewell, William H. and Robert M. Hauser. 1980. "The Wisconsin Longitudinal Study of Social and Psychological Factors in Aspirations and Achievements." In Alan C. Kerckhoff. ed. Research in Sociology of Education and Socialization. Vol. 1. Greenwich: JAI Press. Pp. 59-99.
- Sharma, Prakash C. 1975. A Selected Research Bibliography on Panel Studies. Exchange Bibliography No. 758. Monticello, Illinois: Council of Planning Librarians.
- Shryock, Henry S., Jr. and Elizabeth A. Larmon. 1965. "Some Longitudinal Data on Internal Migration." Demography 2: 579-592.
- Struyk, Raymond J. and Marc Mendick, Jr. 1981. Housing Vouchers for the Poor: Lessons from a National Experiment. Washington, D.C.: Urban Institute.
- U.S. Bureau of the Census. 1982. Wage and Salary Data from the income Survey Development Program: 1979. Current Population Reports, Series P-23, No. 118. Washington, D.C.: U.S. Government Printing Office.
- _____. 1983. Housing Characteristics of Recent Movers. Current Housing Reports, Series H-150-81, Part D. Washington, D.C.: U.S. Government Printing Office.
- U.S. Department of Commerce, Office of Federal Statistical Policy and Standards. 1978. "Longitudinal Surveys." In A Framework for Planning U.S. Federal Statistics for the 1980's. Washington, D.C.: U.S. Government Printing Office. Pp. 321-325.
- U.S. Department of Housing and Urban Development. 1977. The Urban Homesteading Catalogue. 3 volumes. Washington, D.C.: U.S. Government Printing Office.
- U.S. Department of Labor, Bureau of Labor Statistics. 1980. Using the Current Population Survey as a Longitudinal Data Base. Bureau of Labor

Statistics Report No. 608. Washington, D.C.: U.S. Government Printing Office.

Watts, Harold and Albert Rees. eds. 1977. The New Jersey Income-Maintenance Experiment. Vols. 2 and 3. New York: Academic Press.

Wise, Lauress L. and Lauri Steel. 1980. "Educational Attainment of the High School Classes of 1960 Through 1963: Findings from Project TALENT." In

Alan C. Kerckhoff. ed. Research in Sociology of Education and Socialization. Vol. 1. Greenwich: JAI Press. Pp. 101-126.

Wooldridge, Judith. 1977. "Housing Consumption." In Harold W. Watts and Albert Rees. eds. The New Jersey Income-Maintenance Experiment. Vol. 3. New York: Academic Press. Pp. 45-71.

August 15, 1984

SIPP AND CPS LABOR FORCE CONCEPTS:
A COMPARISON

by Paul M. Ryscavage,
Bureau of the Census

To be presented at the 1984 Joint Statistical Meetings, Philadelphia, Pa.,
August 1984.

SIPP AND CPS LABOR FORCE CONCEPTS:
A COMPARISON

by Paul M. Ryscavage

Background

The Survey of Income and Program Participation (SIPP) is a new Census Bureau survey designed to give policymakers, researchers, and the public an in-depth look at the economic situation of persons and households in the United States. Its primary purpose is to collect data on the kinds and amounts of income received by persons and the extent of their participation in government income transfer programs, such as Social Security and Aid to Families with Dependent Children. The full scope of SIPP as a source of information on the well-being of our society, however, is still being realized.

One important byproduct of SIPP is information on the labor force activity of individuals. Working or not working is frequently associated with one's economic situation and also one's participation or nonparticipation in social welfare programs. An obvious illustration is the relationship between job loss and the receipt of unemployment insurance payments.

In the development of the SIPP labor force questions, an effort was made to make them conceptually similar to those in the Current Population Survey (CPS) which is the survey used to collect the Federal government's official labor force statistics. The CPS, in operation since 1940, was developed for the sole purpose of estimating the numbers of employed and unemployed persons in the country.

Paul M. Ryscavage is a labor economist in the Population Division of the Bureau of the Census.

NOTE: In addition to the references footnoted in the text, the SIPP Interviewer's Manual and CPS Interviewers Reference Manual were used in the preparation of this paper.

At the core of the CPS labor force data is the "activity concept." ^{1/} Basically, the concept amounts to identifying persons' activities in relation to the labor market during a specific period of time. In the CPS the period of time is one week. Persons in the adult population can then be sorted into three mutually exclusive groups depending on their activity during the week: working, not working but seeking work, and neither working nor seeking work.

While many refinements have been made to the activity concept and the operation of the CPS through the years, the keystone of the Nation's employment and unemployment estimates -- activity during a specific reference week -- has not been changed. The concept and the CPS have been reviewed periodically by Presidentially appointed commissions to insure their soundness. The most recent review was by the National Commission on Employment and Unemployment Statistics (NCEUS) in the late 1970's. ^{2/} Although the Commission recommended some modifications of definitions used in the CPS, it pronounced the basic activity concept as sound. ^{3/}

Compared to the CPS, SIPP is in its infancy. Its genesis was the Income Survey Development Program begun in the mid-1970's by the Department of Health, Education, and Welfare. ^{4/} Despite its newness, SIPP has great potential for not only casting light on the nature of income dynamics, but also on how labor force

^{1/} For those interested in the origins of the activity concept, see John N. Webb, "Concepts Used In Unemployment Surveys," Journal of the American Statistical Association, March 1939, pp. 49-61. For a history of the CPS, see John E. Bregger, "The Current Population Survey: a historical perspective and BLS' role," Monthly Labor Review, June, 1984, pp. 8-14.

^{2/} See Counting the Labor Force, National Commission on Employment and Unemployment Statistics, (U.S. Government Printing Office) Labor Day, 1979. An earlier review was made in the early 1960's. See Measuring Employment and Unemployment, President's Committee to Appraise Employment and Unemployment Statistics, (U.S. Government Printing Office) 1962.

^{3/} See Counting the Labor Force, p. 2.

^{4/} See Martynas A. Ycas and Charles Lininger, "The Income Survey Development Program: Design Features and Initial Findings," Social Security Bulletin, November 1981, pp. 13-19.

activity is related to it. Indeed, the NCEUS suggested there was a need "to link labor force experience with income data" so as to add a qualitative dimension to labor force statistics.^{5/} SIPP data will show on a regular basis how well the labor market is providing for the economic well-being of workers and their households.

An obvious question among labor force analysts is how will the SIPP and CPS labor force data compare? Although we can't answer that question at this time because SIPP labor force data are still being processed, we can compare SIPP and CPS labor force concepts.^{6/} More specifically, we can examine how the activity concept is applied in both surveys. We begin first by briefly reviewing some of the survey design characteristics of the SIPP and CPS and then compare specific SIPP and CPS labor force definitions. A concluding section of the paper discusses potential uses of SIPP labor force data.

^{5/} See Counting the Labor Force. p. 1.

^{6/} Labor force estimates from the ISDP were compared to the CPS by Bruce Klein. He found that employment estimates in ISDP were slightly higher than in CPS and unemployment estimates in ISDP were considerably lower than in CPS. One reason for the latter was that persons on layoff were not counted in the ISDP. See Bruce W. Klein, "Comparing Labor Force Measures in ISDP with CPS," Technical, Conceptual, and Administrative Lessons of the Income Survey Development Program (ISDP), Papers presented at a Conference (Martin H. David, ed.), Social Science Research Council, New York, New York, 1983, pp. 229-239.

Survey Design Characteristics of SIPP and CPS

Labor force analyses (as well as other kinds of analyses) are frequently limited because the data being analyzed come from surveys with unique survey design characteristics. For example, small sample size often creates difficulties for analysts. Three survey design features of SIPP and CPS which are important from an analytical standpoint are discussed below.

Samples. Significant differences exist in the sample size and design of SIPP and CPS. SIPP is a longitudinal panel survey comprised originally of 26,000 households located in 174 areas around the country. The sample is divided into four rotation groups and households in each group are interviewed every four months for approximately two and one-half years. The first rotation group was interviewed in October 1983, and interviews were conducted in the second, third, and fourth rotation groups in November, December, and January, respectively. This staggered sample design produces a cycle or wave of interviewing and takes four months to complete after which the rotation groups are reinterviewed in the same sequence. The Census Bureau plans to introduce another panel of approximately 20,000 households in January 1985 and another 20,000 household panel in January 1986. Consequently, SIPP's sample size will grow as panels are overlapped, increasing the reliability of the estimates.

The CPS is basically a cross-sectional survey, but it also has a longitudinal dimension. ^{7/} It is a much larger survey being comprised of 60,000 households located in 629 areas across the country. The CPS sample is divided into eight rotation groups but unlike the staggered sample design of SIPP, all rotation groups are in operation in a single month. The longitudinal aspect of CPS results from the rotation group pattern in which a household is in the sample for four

^{7/} For a discussion of the longitudinal nature of the CPS see, Using the Current Population Survey as a Longitudinal Data Base, Report 608, Bureau of Labor Statistics, August 1980.

consecutive months, out for eight and then back in for four more months. This pattern allows three-quarters of the households to be the same from month-to-month and one half to be the same over the year. This is important because labor force analyses of CPS data conducted by the Bureau of Labor Statistics (BLS) concentrate on month-to-month and year-to-year changes.

Two problems for labor force analysts who use household survey data are biases resulting from the sample's design and from interview nonresponse. Rotation group bias has always been a problem in the CPS and it has received much attention over the years. ^{8/} Theoretically, each CPS rotation group should produce the same estimates, except for random differences due to sampling variability. The estimate of unemployment from the first rotation group, however, is usually greater than the estimate based on all rotation groups. ^{9/} (Recently, the difference has averaged about six percent.) The reason for the difference has never been isolated. Because all SIPP rotation groups in a SIPP panel have been in the sample for the same amount of time, this type of bias will not be immediately observable. It will be possible to observe after the introduction of the 1985 SIPP panel in January 1985, however, since then rotation groups of different sample ages will be in operation at the same time.

^{8/} See Barbara Bailar, "The Effects of Rotation Group Bias on Estimates from Panel Surveys," Journal of the American Statistical Association, Vol. 70, March 1975.

^{9/} See the Current Population Survey: Design and Methodology, Technical Paper 40, Bureau of the Census, January 1978, p. 83.

A second bias problem involves survey nonresponse -- unit or total non-response and item nonresponse. In the CPS, the unit noninterview rate has hovered around the four to five percent mark in recent years; item nonresponse rates vary by item, but in the March CPS income questions generally have the highest nonresponse rate.^{10/} The Census Bureau has developed noninterview adjustments and imputation schemes for dealing with these problems. While the first panel in SIPP is less than a year old, it appears that the unit noninterview rate for the first SIPP interviews is about the same as in the CPS. (A cumulative non-interview rate will be available from SIPP as subsequent waves of interviewing is completed.) Item nonresponse in SIPP is presently being investigated at the Census Bureau.^{11/} Because both labor force and income questions are asked at the same time, the quality of the SIPP labor force data may be affected.

Survey eligibility and coverage. Respondent eligibility and coverage are somewhat different in SIPP and CPS. In SIPP all household members 15 years of age and over are eligible to be interviewed and all eligible persons are interviewed if present at the time of the interview. If an eligible person is not home, a "proxy" interview is obtained from a knowledgeable person, otherwise a return visit is scheduled. In the CPS the age of eligibility is 16 years and over (data are also collected for 14 and 15 year olds); one adult household respondent may answer the questions for all household members.

Telephone interviewing is also handled differently in the two surveys. Telephone interviews in SIPP must have prior regional office approval, except in the case of information not obtained in the course of the interview. In the the CPS, telephone interviews are permitted in the second, third, fourth, sixth,

^{10/} Ibid., p. 87.

^{11/} See John Coder and Angela Feldman, "Early Indications of Item Nonresponse on the Survey of Income and Program Participation," a paper to be presented at the 1984 Joint Statistical Meetings, Philadelphia, Pa., August 1984.

seventh, and eighth month in which the households are in sample.

Another difference concerns the treatment of the Armed Forces. In the monthly CPS, members of the Armed Forces living in households are not eligible for interview. In SIPP, however, such individuals are interviewed as long as they are stationed in the area and usually reside at the address visited. (Both surveys exclude inmates of institutions, such as persons in prisons or convalescent homes.)

Lastly, and most significant for many analyses, members of households in the SIPP sample who move between interviews are followed and further interviews attempted. Sample persons, however, are not followed when they have been institutionalized, become a member of the Armed Forces, move outside the United States, or move more than 100 miles from a SIPP sampling area. In the CPS, movers are not followed and this has been a constraint on many longitudinal labor force analyses. ^{12/}

Reference periods. A fundamental difference between SIPP and CPS -- one that will probably account for differences in labor force estimates between surveys -- is the length of the reference period. CPS interviews are conducted in all rotation groups each month in the week containing the 19th and all questions about labor force activity are asked in reference to the previous week which contains the 12th, the survey week. (As will be discussed, this one week reference period is extended to four weeks in the case of jobseeking.) Depending on the respondent's answers to the questions, household members are classified into one of three

^{12/} See, for example, Francis W. Horvath, "Tracking Individual Earnings Mobility with the Current Population Survey," Monthly Labor Review, May 1980, pp. 43-46.

mutually exclusive groups, employed, unemployed, or not in the labor force.

(See Figure 1.)

SIPP interviews are conducted in one of the four rotation groups each month during the first two weeks of the month. The labor force, income, and program participation questions relate to the four previous months. Indeed, the labor force questions actually refer to individual weeks during the four month period. During this time a person could have worked, looked for work, and been outside the labor force at different times. In other words, labor force classification in SIPP is not necessarily mutually exclusive as it is in the CPS.^{13/} (See Figure 1.)

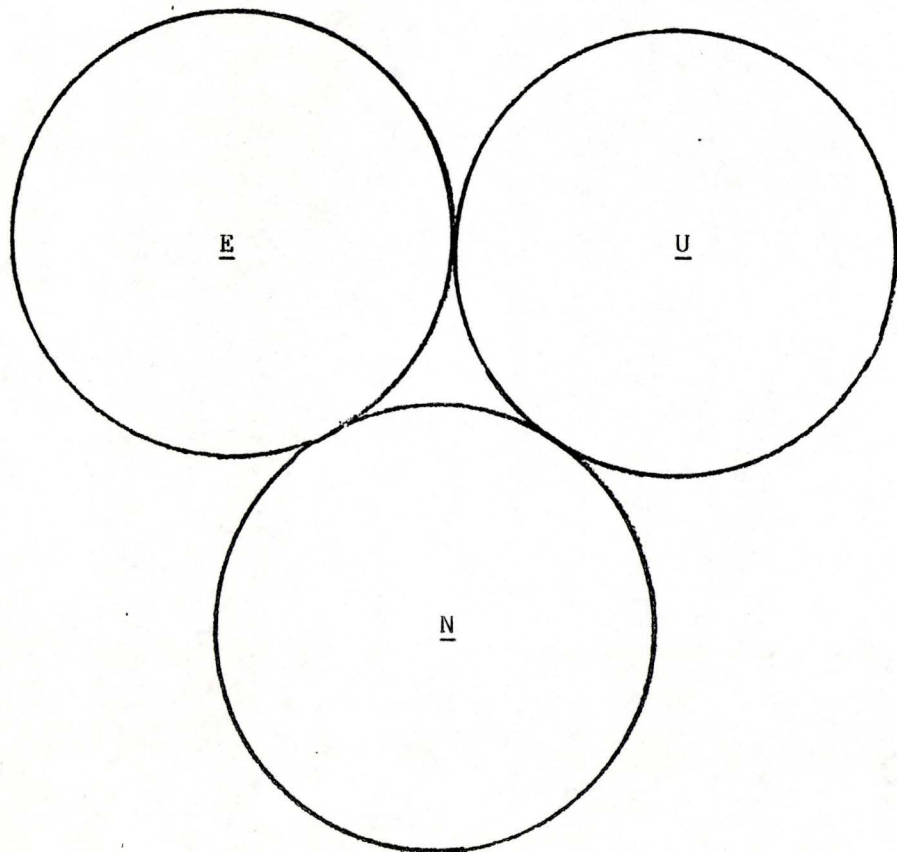
Recall problems are potentially a greater problem in SIPP than in the monthly CPS since respondents are recalling activities over a much longer period. For persons with a marginal attachment to the labor market, for example, teenagers, it may be very difficult to remember job market activities. Despite the long recall period in SIPP, it is not inordinately long. In the supplement to the March CPS persons are asked about their labor market activities in the previous calendar year -- a reference period extending back 3 to 15 months.^{14/} The annual work experience statistics have been published by the BLS and Census Bureau for years.

^{13/} A similar situation prevails in the March supplement to the CPS where persons are asked about their work experience in the previous calendar year.

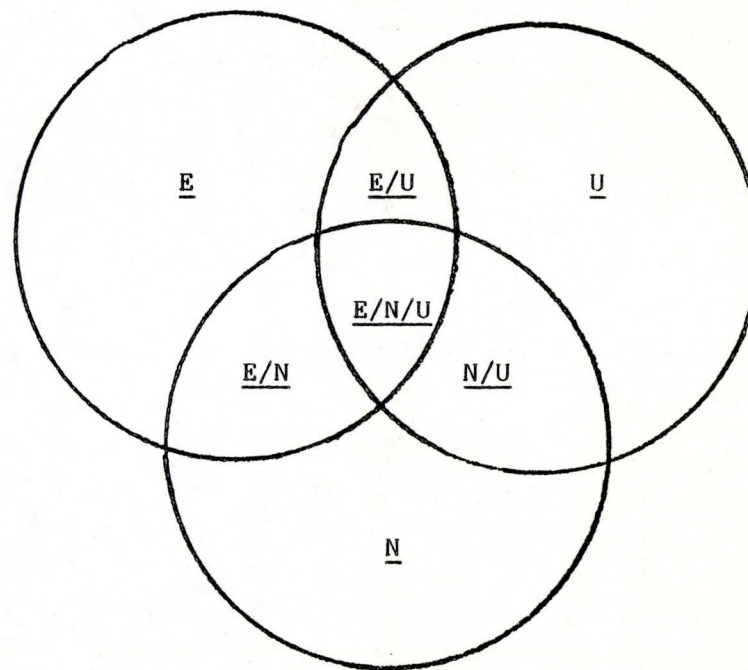
^{14/} The retrospective bias in the March CPS work experience data has been the topic of research in recent years. For example, see Richard D. Morgenstern and Nancy S. Barrett. "The Retrospective Bias in Unemployment Reporting by Sex, Race, and Age," Journal of the American Statistical Association, June 1974, pp. 355-357. For more recent research see, Francis W. Horvath, "Forgotten Unemployment: Recall Bias in Retrospective Data," Monthly Labor Review, March 1982, pp. 40-43.

FIGURE 1. Possible Labor Force Status Categories During CPS and SIPP Reference Periods

CPS (one week)



SIPP (four months)



E - Employed
U - Unemployed
N - Not in labor force

SIPP and CPS Labor Force Definitions

Because the reference periods in SIPP and CPS are of different lengths, the activity concept is applied differently in both surveys. In the CPS, persons are asked a specific activity-type question relating to the week containing the 12th of the month. (See CPS questionnaire.) In sorting out the possible labor market-related activities into mutually exclusive groups, a priority scheme is necessary since some individuals may have been involved in more than one activity. The first or highest priority is assigned to working. As long as a person worked for pay or profit for one hour or more (or 15 hours or more without pay in a family operated business), the person is considered employed even though he or she may also have looked for work or gone to school or done something else during that week.

The next highest priority is given to those persons who had a job during the survey week, but were temporarily absent from it. Although this relaxes the activity concept slightly, it permits a more accurate counting of the numbers of persons with actual job commitments. These persons may have been on strike or ill or on vacation or absent for some other personal reason, but since they had a job to return to they are classified as employed.

The third priority is assigned to persons whose activity was looking for work. If a person in the survey week neither worked nor had a job but looked for one at some time within the past four weeks (and was currently available to take one) he or she is considered unemployed. Once again the activity concept is relaxed to cover persons who may not have looked for work continuously because they were waiting to be recalled from layoff or were waiting to start a new wage or salary job within 30 days. These persons too would be classified as unemployed. All other individuals not fitting into this classification scheme

CURRENT POPULATION SURVEY

<p>18. LINE NUMBER</p>	<p>19. What was ... doing most of LAST WEEK -</p> <p>Working <input checked="" type="checkbox"/> 2</p> <p>Keeping house <input type="checkbox"/> 3</p> <p>Going to school or something else? <input type="checkbox"/> 4</p> <p>Working (Ship to 20A) <input type="checkbox"/> 5</p> <p>With a job but not at work <input type="checkbox"/> 6</p> <p>Looking for work <input type="checkbox"/> 7</p> <p>Keeping house <input type="checkbox"/> 8</p> <p>Going to school <input type="checkbox"/> 9</p> <p>Unable to work (Ship to 24) <input type="checkbox"/> 0</p> <p>Retired <input type="checkbox"/> 1</p> <p>Other (Specify) <input type="checkbox"/> 2</p>	<p>20. Did ... go any work at all LAST WEEK, not counting work around the house?</p> <p>(Note: If farm or business operator in hhs, ask about unpaid work.)</p> <p>Yes <input checked="" type="checkbox"/> 2 No <input type="checkbox"/> 1 (Go to 21)</p> <p>20A. How many hours did ... work LAST WEEK at all jobs?</p> <p>0 <input type="checkbox"/> 0 1 <input type="checkbox"/> 1 2 <input type="checkbox"/> 2 3 <input type="checkbox"/> 3 4 <input type="checkbox"/> 4 5 <input type="checkbox"/> 5 6 <input type="checkbox"/> 6 7 <input type="checkbox"/> 7 8 <input type="checkbox"/> 8 9 <input type="checkbox"/> 9</p> <p>20B. INTERVIEWER CHECK ITEM</p> <p>49 <input type="checkbox"/> (Ship to item 23)</p> <p>1-34 <input type="checkbox"/> (Go to 20C)</p> <p>35-48 <input type="checkbox"/> (Go to 20D)</p> <p>20C. Did ... lose any time or take any time off LAST WEEK for any reason such as illness, holiday or slack work?</p> <p>Yes <input type="checkbox"/> How many hours did ... take off?</p> <p>(Correct 20A if last time not already deducted; if 20A reduced below 35, correct 20B and fill 20C; otherwise, skip to 23.)</p> <p>No <input checked="" type="checkbox"/> 7</p> <p>20D. Did ... work any overtime or at more than one job LAST WEEK?</p> <p>Yes <input type="checkbox"/> How many extra hours did ... work?</p> <p>(Correct 20A and 20B as necessary if extra hours not already included and skip to 23.)</p> <p>No <input checked="" type="checkbox"/> 8 (Ship to 23)</p>	<p>21. (If in 19, skip to 21A.)</p> <p>Did ... have a job or business from which he/she was temporarily absent or on layoff LAST WEEK?</p> <p>Yes <input type="checkbox"/> 0 No <input checked="" type="checkbox"/> 1 (Go to 22)</p> <p>21A. Why was ... absent from work LAST WEEK?</p> <p>Own illness <input type="checkbox"/> 0</p> <p>On vacation <input type="checkbox"/> 1</p> <p>Bad weather <input type="checkbox"/> 2</p> <p>Labor dispute <input type="checkbox"/> 3</p> <p>New job to begin within 30 days (Ship to 22B and 22C)</p> <p>Temporary layoff (Under 30 days)</p> <p>Indefinite layoff (30 days or more or no est. recall date) (Ship to 22C)</p> <p>Other (Specify) ...</p> <p>21B. Is ... getting wages or salary for any of the time off LAST WEEK?</p> <p>Yes <input type="checkbox"/> 0 No <input checked="" type="checkbox"/> 1</p> <p>Self-employed <input type="checkbox"/> 2</p> <p>21C. Does ... usually work 35 hours or more a week at this job?</p> <p>Yes <input type="checkbox"/> 0 No <input checked="" type="checkbox"/> 1 (Ship to 23 and enter job held last week)</p>	<p>22. (If LK in 19, skip to 22A.)</p> <p>Has ... been looking for work during the past 4 weeks?</p> <p>Yes <input checked="" type="checkbox"/> 2 No <input type="checkbox"/> 1 (Go to 24)</p> <p>22A. What has ... been doing in the last 4 weeks to find work? (Mark all methods used; do not read list.)</p> <p>Checked with -</p> <p>pub. emp. agency <input type="checkbox"/> 0</p> <p>priv. emp. agency <input type="checkbox"/> 1</p> <p>employer directly <input type="checkbox"/> 2</p> <p>friends or relatives <input type="checkbox"/> 3</p> <p>Placed or answered ads <input type="checkbox"/> 4</p> <p>Nothing (Ship to 24) <input type="checkbox"/> 5</p> <p>Other (Specify in notes, e.g., CETA, union or prof. register, etc.) <input type="checkbox"/> 6</p> <p>22B. Why did ... start looking for work? Was it because ... lost or quit a job at that time (paste or was there some other reason?)</p> <p>Last job <input type="checkbox"/> 0</p> <p>Quit job <input type="checkbox"/> 1</p> <p>Laid school <input type="checkbox"/> 2</p> <p>Wanted temporary work <input type="checkbox"/> 3</p> <p>Other (Specify in notes) <input type="checkbox"/> 4</p> <p>22C. 1) How many weeks has ... been looking for work?</p> <p>0 <input type="checkbox"/> 0 1 <input type="checkbox"/> 1 2 <input type="checkbox"/> 2 3 <input type="checkbox"/> 3 4 <input type="checkbox"/> 4 5 <input type="checkbox"/> 5 6 <input type="checkbox"/> 6 7 <input type="checkbox"/> 7 8 <input type="checkbox"/> 8 9 <input type="checkbox"/> 9</p> <p>2) How many weeks ago did ... start looking for work?</p> <p>0 <input type="checkbox"/> 0 1 <input type="checkbox"/> 1 2 <input type="checkbox"/> 2 3 <input type="checkbox"/> 3 4 <input type="checkbox"/> 4 5 <input type="checkbox"/> 5 6 <input type="checkbox"/> 6 7 <input type="checkbox"/> 7 8 <input type="checkbox"/> 8 9 <input type="checkbox"/> 9</p> <p>3) How many weeks ago was ... laid off?</p> <p>0 <input type="checkbox"/> 0 1 <input type="checkbox"/> 1 2 <input type="checkbox"/> 2 3 <input type="checkbox"/> 3 4 <input type="checkbox"/> 4 5 <input type="checkbox"/> 5 6 <input type="checkbox"/> 6 7 <input type="checkbox"/> 7 8 <input type="checkbox"/> 8 9 <input type="checkbox"/> 9</p> <p>22D. Has ... been looking for full-time or part-time work?</p> <p>Full <input type="checkbox"/> 0 Part <input checked="" type="checkbox"/> 1</p> <p>22E. Is there any reason why ... could not take a job LAST WEEK?</p> <p>Yes <input type="checkbox"/> 0 No <input checked="" type="checkbox"/> 1</p> <p>Already has a job <input type="checkbox"/> 0</p> <p>Temporary illness <input type="checkbox"/> 1</p> <p>Going to school <input type="checkbox"/> 2</p> <p>Other (Specify in notes) <input type="checkbox"/> 3</p> <p>22F. When did ... last work at a full-time job or business lasting 2 consecutive weeks or more? Within last 12 months (Specify) ...</p> <p>One to five years ago <input type="checkbox"/> 0</p> <p>More than 5 years ago <input type="checkbox"/> 1</p> <p>Never worked full-time 2 wks. or more <input type="checkbox"/> 2</p> <p>Never worked at all <input type="checkbox"/> 3</p> <p>(SKIP to 23. If layoff entered in 21A, enter job, either full or part time, from which laid off. Else enter last full time job lasting 2 weeks or more, or "never worked.")</p>	<p>24. INTERVIEWER CHECK ITEM</p> <p>(Rotation number)</p> <p>First digit of SEGMENT number is:</p> <p>0 1, 2, 3, 5, 6 or 7 (End questions)</p> <p>4 or 8 (Go to 24A)</p> <p>24A. When did ... last work for pay at a regular job or business, either full- or part-time?</p> <p>Within past 12 months <input checked="" type="checkbox"/> 2</p> <p>1 up to 2 years ago <input type="checkbox"/> 1</p> <p>2 up to 3 years ago <input type="checkbox"/> 0 (Go to 24B)</p> <p>3 up to 4 years ago <input type="checkbox"/> 9</p> <p>4 up to 5 years ago <input type="checkbox"/> 8</p> <p>5 or more years ago <input type="checkbox"/> 7</p> <p>Never worked <input type="checkbox"/> 6 (Ship to 24C)</p> <p>24B. Why did ... leave that job?</p> <p>Personal, family (incl. pregnancy) or school <input type="checkbox"/> 0</p> <p>Health <input type="checkbox"/> 1</p> <p>Retirement or old age <input type="checkbox"/> 2</p> <p>Seasonal job completed <input type="checkbox"/> 3</p> <p>Slack work or business conditions <input type="checkbox"/> 4</p> <p>Temporary nonseasonal job completed <input type="checkbox"/> 5</p> <p>Unsatisfactory work arrangements (Hours, pay, etc.) <input type="checkbox"/> 6</p> <p>Other <input type="checkbox"/> 7</p> <p>24C. Does ... want a regular job now, either full- or part-time?</p> <p>Yes <input checked="" type="checkbox"/> 2 Maybe - it depends <input type="checkbox"/> 1 (Go to 24D)</p> <p>No <input type="checkbox"/> 0 Don't know <input type="checkbox"/> 3 (Ship to 24E)</p> <p>24D. What are the reasons ... is not looking for work? (Mark each reason mentioned)</p> <p>Settles no work available in line of work or area <input type="checkbox"/> 0</p> <p>Couldn't find any work <input type="checkbox"/> 1</p> <p>Lacks nec. schooling, training, skills or experience <input type="checkbox"/> 2</p> <p>Employers think too young or too old <input type="checkbox"/> 3</p> <p>Other pers. handicap in finding job <input type="checkbox"/> 4</p> <p>Can't arrange child care <input type="checkbox"/> 5</p> <p>Family responsibilities <input type="checkbox"/> 6</p> <p>In school or other training <input type="checkbox"/> 7</p> <p>Ill health, physical disability <input type="checkbox"/> 8</p> <p>Other (Specify in notes) <input type="checkbox"/> 9</p> <p>Don't know <input type="checkbox"/> 0</p> <p>24E. Does ... intend to look for work of any kind in the next 12 months?</p> <p>Yes <input type="checkbox"/> 0 It depends (Specify in notes) <input type="checkbox"/> 1</p> <p>No <input type="checkbox"/> 2 Don't know <input type="checkbox"/> 3 (If entry in 24B, describe job in 23, otherwise, end questions.)</p>	<p>25. INTERVIEWER CHECK ITEM</p> <p>(Rotation number)</p> <p>First digit of SEGMENT number is:</p> <p>0 1, 2, 3, 5, 6 or 7 (End questions)</p> <p>4 or 8 (Go to 25A)</p> <p>25A. How many hours per week does ... USUALLY work at this job?</p> <p>0 <input type="checkbox"/> 0 1 <input type="checkbox"/> 1 2 <input type="checkbox"/> 2 3 <input type="checkbox"/> 3 4 <input type="checkbox"/> 4 5 <input type="checkbox"/> 5 6 <input type="checkbox"/> 6 7 <input type="checkbox"/> 7 8 <input type="checkbox"/> 8 9 <input type="checkbox"/> 9</p> <p>25B. Is ... paid by the hour on this job?</p> <p>Yes <input type="checkbox"/> (Go to 25C) No <input checked="" type="checkbox"/> 1 (Ship to 25D)</p> <p>25C. How much does ... earn per hour?</p> <table border="1" style="width: 100%; text-align: center;"> <tr> <td>Dollars</td> <td>Cents</td> </tr> <tr> <td>0 0 0 0</td> <td>0 0</td> </tr> <tr> <td>1 0 0 0</td> <td>0 0</td> </tr> <tr> <td>2 0 0 0</td> <td>0 0</td> </tr> <tr> <td>3 0 0 0</td> <td>0 0</td> </tr> <tr> <td>4 0 0 0</td> <td>0 0</td> </tr> <tr> <td>5 0 0 0</td> <td>0 0</td> </tr> <tr> <td>6 0 0 0</td> <td>0 0</td> </tr> <tr> <td>7 0 0 0</td> <td>0 0</td> </tr> <tr> <td>8 0 0 0</td> <td>0 0</td> </tr> <tr> <td>9 0 0 0</td> <td>0 0</td> </tr> </table> <p>(Ask 25D)</p> <p>25D. How much does ... USUALLY earn per week at this job BEFORE deductions? Include any overtime pay, commissions, or tips usually received.</p> <table border="1" style="width: 100%; text-align: center;"> <tr> <td>Dollars</td> <td>Cents</td> </tr> <tr> <td>0 0 0 0</td> <td>0 0</td> </tr> <tr> <td>1 0 0 0</td> <td>0 0</td> </tr> <tr> <td>2 0 0 0</td> <td>0 0</td> </tr> <tr> <td>3 0 0 0</td> <td>0 0</td> </tr> <tr> <td>4 0 0 0</td> <td>0 0</td> </tr> <tr> <td>5 0 0 0</td> <td>0 0</td> </tr> <tr> <td>6 0 0 0</td> <td>0 0</td> </tr> <tr> <td>7 0 0 0</td> <td>0 0</td> </tr> <tr> <td>8 0 0 0</td> <td>0 0</td> </tr> <tr> <td>9 0 0 0</td> <td>0 0</td> </tr> </table> <p>25E. On this job, is ... a member of a labor union or of an employees association similar to a union?</p> <p>Yes <input type="checkbox"/> (End questions) No <input checked="" type="checkbox"/> 1 (Ask 25F)</p> <p>25F. On this job, is ... covered by a union or employees association contract?</p> <p>Yes <input type="checkbox"/> (End questions) No <input checked="" type="checkbox"/> 1 (End questions)</p>	Dollars	Cents	0 0 0 0	0 0	1 0 0 0	0 0	2 0 0 0	0 0	3 0 0 0	0 0	4 0 0 0	0 0	5 0 0 0	0 0	6 0 0 0	0 0	7 0 0 0	0 0	8 0 0 0	0 0	9 0 0 0	0 0	Dollars	Cents	0 0 0 0	0 0	1 0 0 0	0 0	2 0 0 0	0 0	3 0 0 0	0 0	4 0 0 0	0 0	5 0 0 0	0 0	6 0 0 0	0 0	7 0 0 0	0 0	8 0 0 0	0 0	9 0 0 0	0 0
Dollars	Cents																																																	
0 0 0 0	0 0																																																	
1 0 0 0	0 0																																																	
2 0 0 0	0 0																																																	
3 0 0 0	0 0																																																	
4 0 0 0	0 0																																																	
5 0 0 0	0 0																																																	
6 0 0 0	0 0																																																	
7 0 0 0	0 0																																																	
8 0 0 0	0 0																																																	
9 0 0 0	0 0																																																	
Dollars	Cents																																																	
0 0 0 0	0 0																																																	
1 0 0 0	0 0																																																	
2 0 0 0	0 0																																																	
3 0 0 0	0 0																																																	
4 0 0 0	0 0																																																	
5 0 0 0	0 0																																																	
6 0 0 0	0 0																																																	
7 0 0 0	0 0																																																	
8 0 0 0	0 0																																																	
9 0 0 0	0 0																																																	
<p>23. DESCRIPTION OF JOB OR BUSINESS</p>																																																		
<p>23A. For whom did ... work? (Name of company, business, organization or other employer.)</p>	<p>23B. What kind of business or industry is this? (For example: TV and radio mfg., retail shoe store, State Labor Dept., farm.)</p>	<p>23C. What kind of work was ... doing? (For example: electrical engineer, stock clerk, typist, farmer.)</p>																																																
<p>23D. What were ...'s most important activities or duties at the job? (For example: types, Aves account books, files, sells cars, operates printing press, finishes concrete.)</p>	<p>23E. Was this person an employee of PRIVATE Co., bus., or individual for wages, salary or comm. ... P <input type="checkbox"/></p> <p>A FEDERAL government employee ... F <input type="checkbox"/> (Go to 23F)</p> <p>A STATE government employee ... S <input type="checkbox"/></p> <p>A LOCAL government employee ... L <input type="checkbox"/></p> <p>Self-empl. in OWN bus., prof. practice, or farm <input checked="" type="checkbox"/> 2</p> <p>Is the business incorporated? Yes <input type="checkbox"/> 1 No <input checked="" type="checkbox"/> 0</p> <p>Working WITHOUT PAY in fam. bus. or farm ... WP <input type="checkbox"/> (End questions)</p> <p>NEVER WORKED ... NEV <input type="checkbox"/> (End questions)</p>																																																	
<p>23F. INTERVIEWER CHECK ITEM</p> <p>Entry for NAJ in item 23A <input type="checkbox"/> (Go to 25 at top of Page)</p> <p>Entry for NAJ in item 21B <input type="checkbox"/> (End questions)</p> <p>All other cases <input type="checkbox"/> (End questions)</p>																																																		

are considered to be not in the labor force. Accordingly, the Nation's civilian noninstitutional population age 16 and over can be sorted into the familiar labor force categories shown below.

CPS Labor Force Categories

Civilian noninstitutional population age 16 and over	
Civilian labor force	
Employed	
Unemployed	
Not in the labor force	

In SIPP persons are not asked a specific activity question relating to the previous four months because it is very possible (even more so than in the CPS) that an individual may have worked, looked for work, or done something else during the period. Instead, the initial question on the SIPP questionnaire concerns whether or not an individual had a job or business at any time during the previous four months. (See the SIPP questionnaire.) In other words, the activity concept is tied to an individual's either having or not having a job in the reference period. For those who had jobs, subsequent questions are asked about how long persons had their jobs in the reference period, whether they had been absent from them and why, and whether they looked for work or were on layoff when they did not have jobs. For persons who did not have jobs during the entire period, questions are asked if they looked for work or had been on layoff and if so, for how long.

Unlike the CPS where a priority scheme is required to classify individuals into mutually exclusive labor force categories, in SIPP individuals may have experienced more than one labor force status in the four month reference period. For example, a person may have had a job for the entire period but was temporarily laid off for one month of the period; or a person may have had a job for two

SURVEY OF INCOME AND PROGRAM PARTICIPATION
(labor force and employment section)

LABOR FORCE AND RECEIENCY

Section 1 – LABOR FORCE AND RECEIENCY

(SHOW FLASHCARD J)

PGM 7

1. During the 4-month period outlined on this calendar, that is, from (4 months ago) thru (Last month), did ... have a job or business, either full time or part time, even for only a few days?
Mark "Yes" for active duty in the Armed Forces, any temporary or part-time work, and work without pay in a family business or farm.

- 1000 1 Yes – Mark "Worked" (code 170) on ISS and SKIP to 4
2 No

2a. Even though ... did not have a job during this period, did ... spend any time looking for work or on layoff from a job?

- 1002 1 Yes
2 No – SKIP to 3a

b. Please look at the calendar. In which weeks was ... looking for work or on layoff from a job?

Mark (X) all that apply.

- 1004 xs ALL
- | | | |
|---------------------------------|----------------------------------|----------------------------------|
| 1006 <input type="checkbox"/> 1 | 1018 <input type="checkbox"/> 7 | 1030 <input type="checkbox"/> 13 |
| 1008 <input type="checkbox"/> 2 | 1020 <input type="checkbox"/> 8 | 1032 <input type="checkbox"/> 14 |
| 1010 <input type="checkbox"/> 3 | 1022 <input type="checkbox"/> 9 | 1034 <input type="checkbox"/> 15 |
| 1012 <input type="checkbox"/> 4 | 1024 <input type="checkbox"/> 10 | 1036 <input type="checkbox"/> 16 |
| 1014 <input type="checkbox"/> 5 | 1026 <input type="checkbox"/> 11 | 1038 <input type="checkbox"/> 17 |
| 1016 <input type="checkbox"/> 6 | 1028 <input type="checkbox"/> 12 | 1040 <input type="checkbox"/> 18 |

c. Could ... have taken a job during any of those weeks if one had been offered?

- 1042 1 Yes – SKIP to Check Item R1
2 No

d. What was the main reason ... could not take a job during those weeks?

Mark (X) only one.

- 1044 1 Already had a job
2 Temporary illness
3 School
4 Other – Specify _____

CHECK ITEM 11

Refer to item 2b.
Is the "ALL" box marked in 2b?

- 1046 1 Yes – SKIP to 9a, page 4
2 No – SKIP to 3b

3a. Were there any weeks in the 4-month period when ... wanted a job?

- 1048 1 Yes – SKIP to 3c
2 No – SKIP to Check Item R6; page 4

b. I have recorded that there were weeks that ... did not work or look for work. Did ... want a job in those weeks?

- 1050 1 Yes
2 No – SKIP to 9a, page 4

c. Could ... have taken a job in those weeks if one had been offered?

- 1052 1 Yes
2 No – SKIP to 9a, page 4

d. During the weeks that ... wanted a job but was not looking for one, what was the main reason ... was not looking?

Mark (X) only one.

- 1054 1 Believes no work available in line of work or area
2 Couldn't find any work
3 Lacks necessary schooling, training, skills, or experience
4 Employers think too young or too old
5 Other personal handicap in finding job
6 Can't arrange child care
7 Family responsibilities
8 In school or other training
9 Ill health, physical disability
10 Other – Specify _____
x1 DK

SKIP to 9a, page 4

4. Did ... have a job or business, either full or part time, during EACH of the weeks in this period?
Note that the person did not have to work each week.

- 1056 1 Yes
2 No – SKIP to 6a

5a. Was ... absent without pay from ...'s job or business for any FULL weeks during the 4-month period?

- 1058 1 Yes
2 No – SKIP to 8a, page 4

b. Please look at the calendar. In which weeks was ... absent without pay?

Mark (X) all that apply.

- 1060 xs ALL
- | | | |
|---------------------------------|----------------------------------|----------------------------------|
| 1062 <input type="checkbox"/> 1 | 1074 <input type="checkbox"/> 7 | 1086 <input type="checkbox"/> 13 |
| 1064 <input type="checkbox"/> 2 | 1076 <input type="checkbox"/> 8 | 1088 <input type="checkbox"/> 14 |
| 1066 <input type="checkbox"/> 3 | 1078 <input type="checkbox"/> 9 | 1090 <input type="checkbox"/> 15 |
| 1068 <input type="checkbox"/> 4 | 1080 <input type="checkbox"/> 10 | 1092 <input type="checkbox"/> 16 |
| 1070 <input type="checkbox"/> 5 | 1082 <input type="checkbox"/> 11 | 1094 <input type="checkbox"/> 17 |
| 1072 <input type="checkbox"/> 6 | 1084 <input type="checkbox"/> 12 | 1096 <input type="checkbox"/> 18 |

c. What was the main reason ... was absent from ...'s job or business during those weeks?

Mark (X) only one.

- 1098 1 On layoff
2 Own illness
3 On vacation
4 Bad weather
5 Labor dispute
6 New job to begin within 30 days
7 Other – Specify _____

SKIP to 8a, page 4

Section 1 - LABOR FORCE AND RECEIENCY (Continued)

7h. During the weeks that ... wanted a job but was not looking for one, what was the main reason ... was not looking?

Mark (X) only one.

- 1228
- 1 Believes no work available in line of work or area
 - 2 Couldn't find any work
 - 3 Lacks necessary schooling, training, skills, or experience
 - 4 Employers think too young or too old
 - 5 Other personal handicap in finding job
 - 6 Can't arrange child care
 - 7 Family responsibilities
 - 8 In school or other training
 - 9 Ill health, physical disability
 - 10 Other - Specify _____
 - x1 DK

8a. In the weeks that ... worked during the 4-month period, how many hours did ... usually work per week?

- 1230
- Hours per week
- x3 None
 - x1 DK
- } SKIP to Check Item R4

CHECK ATTENTION

Refer to item 8a.
Did ... usually work 35 or more hours per week?

- 1232
- 1 Yes
 - 2 No - SKIP to 8c

8b. Did ... work fewer than 35 hours in any of the weeks that ... worked during this period? Exclude time off WITH PAY because of holidays, vacation, days off or sickness.

- 1234
- 1 Yes
 - 2 No - SKIP to Check Item R4

c. In how many weeks did ... work fewer than 35 hours during this 4-month period?

- 1236
- x5 All
- Weeks

d. What was the main reason ... worked fewer than 35 hours in those weeks?

Mark (X) only one.

- 1238
- 1 Could not find a full-time job
 - 2 Wanted to work part time
 - 3 Health condition or disability
 - 4 Normal working hours are fewer than 35 hours
 - 5 Slack work or material shortage
 - 6 Other - Specify _____

Section 2 – EARNINGS AND EMPLOYMENT (Continued)	
Part A1 – EMPLOYER IDENTIFICATION NUMBER 1	
<p>2a. What is the name of the employer for whom ... worked during this 4-month period?</p> <p><i>(If ... worked for more than one employer, enter the employer for whom ... worked the most hours during the 4-month period or the most recent employer.)</i></p>	<p>PGM 8 Employer Name</p> <p>2000 _____</p>
<p>2b. What kind of business or industry was (Name of company or business)?</p> <p><i>For example: TV and radio manufacturing, retail shoe store, State Labor Department, farm.</i></p>	<p>PGM 8 Employer I.D. No.</p> <p>2002 <input type="checkbox"/></p> <p>PGM 8 _____</p> <p>2004 _____</p>
<p>c. ASK OR VERIFY –</p> <p>Is it mainly –</p>	<p>PGM 8</p> <p>2006</p> <p>1 <input type="checkbox"/> Manufacturing?</p> <p>2 <input type="checkbox"/> Wholesale Trade?</p> <p>3 <input type="checkbox"/> Retail Trade?</p> <p>4 <input type="checkbox"/> Some other kind of business?</p>
<p>d. What kind of work was ... doing on this job?</p> <p><i>For example: Electrical engineer, stock clerk, typist, farmer</i></p>	<p>PGM 8</p> <p>2008 _____</p>
<p>e. What were ...'s main activities or duties?</p> <p><i>For example: Types, keeps account books, files, sells cars, operates printing press, finishes concrete.</i></p>	<p>PGM 8</p> <p>2010 _____</p>
<p>f. ASK OR VERIFY –</p> <p>Was ... an employee of –</p>	<p>PGM 8</p> <p>2012</p> <p>1 <input type="checkbox"/> A private company or individual?</p> <p>2 <input type="checkbox"/> Federal government (exclude Armed Forces)?</p> <p>3 <input type="checkbox"/> State government?</p> <p>4 <input type="checkbox"/> Local government?</p> <p>5 <input type="checkbox"/> Armed Forces?</p> <p>6 <input type="checkbox"/> Unpaid in family business or farm? – SKIP to Check Item E5</p>
<p>3a. ASK OR VERIFY –</p> <p>Was ... employed by (Name of employer) during the entire 4-month period?</p>	<p>PGM 7</p> <p>2014</p> <p>1 <input type="checkbox"/> Yes – SKIP to 4</p> <p>2 <input type="checkbox"/> No</p>
<p>b. When was ... employed by (Name of employer) during this 4-month period?</p>	<p>FROM</p> <p>2018 <input type="text"/> <input type="text"/> Month 2018 <input type="text"/> <input type="text"/> Day</p> <p>TO</p> <p>2020 <input type="text"/> <input type="text"/> Month 2022 <input type="text"/> <input type="text"/> Day</p>
<p>4. ASK OR VERIFY –</p> <p>How many hours per week did ... usually work at this job?</p>	<p>2024 <input type="text"/> <input type="text"/> Hours</p> <p>x3 <input type="checkbox"/> None</p> <p>x1 <input type="checkbox"/> DK</p>
<p>5. Was ... paid by the hour on this job?</p>	<p>2028</p> <p>1 <input type="checkbox"/> Yes</p> <p>2 <input type="checkbox"/> No – SKIP to 7</p>
<p>6. What was ...'s regular hourly pay rate at the end of (Read last month or "to" date in item 3b)?</p>	<p>2028 \$ <input type="text"/> <input type="text"/> . <input type="text"/></p> <p>x1 <input type="checkbox"/> DK</p> <p>x2 <input type="checkbox"/> Ref. – SKIP to Check Item E5</p>
<p>7. During the 4-month period how often was ... paid on this job?</p>	<p>2030</p> <p>1 <input type="checkbox"/> Once a week</p> <p>2 <input type="checkbox"/> Once each 2 weeks</p> <p>3 <input type="checkbox"/> Once a month</p> <p>4 <input type="checkbox"/> Twice a month</p> <p>5 <input type="checkbox"/> Some other way – Specify</p>

months and then quit to look for another one; or a person may have been outside the labor force for a month, looked for a job for a month and then, having found one, worked for two months. Consequently, the SIPP labor force categories shown below reflect multiple labor force statuses.

SIPP Labor Force Categories

Noninstitutional population age 16 and over	
Persons with some labor force activity in period	
With a job the entire period	
Worked all weeks	
Missed some weeks	
Spent time on layoff	
With a job during part of the period	
Spent time looking for work or on layoff	
No job during the entire period	
Spent time looking for work or on layoff entire period	
Spent time looking for work or on layoff some weeks	
Persons with no labor force activity in period	

A closer look at specific SIPP and CPS labor definitions is presented below and key differences in the two surveys are displayed in Table 1. (The definitions are discussed under headings common in everyday usage and should not be construed as CPS-specific labor force terminology.)

Employment. In both surveys, employment is generally defined as working at a job or business for pay or profit at some time in the reference period. A job is considered to be an arrangement for regular work for pay where payment is in cash wages or salaries, at piece rates, in tips, by commission, or in-kind (meals, living quarters, supplies received). A business is defined as an activity which involves the use of machinery or equipment in which money has been invested or an activity requiring an office or "place of business" or an activity which requires advertising. Payment may be in the form of profits or fees. Both surveys also consider persons to be employed when they have

TABLE 1. Differences and Similarities in SIPP and CPS Labor Force Concepts

	<u>SIPP</u>	<u>CPS (monthly)</u>
<u>Survey Design Characteristics</u>		
Sample size	25,000 households	60,000 households
Sample areas	174 areas	629 areas
Interview eligibility	Age 15 and over	Age 14 and over
Type of interview	Interviews attempted for all	One person answers for all
Persons who move	Followed	Not followed
Reference period	Previous 4 months	Previous week
<u>Labor Force Definitions</u>		
Employed	<p>--Persons who worked at a job or business in last 4 months.</p> <p>--Persons with a job or business in last 4 months who did not work because of illness, vacation, bad weather, labor dispute, or personal reasons.</p>	<p>--Persons who worked at a job or business in reference week.</p> <p>--Persons with a job or business in reference week who did not work because of illness, vacation, bad weather, labor dispute, or personal reasons.</p>
Unemployed	<p>--Persons who were jobless, available to work, and looking for work or on layoff during the last 4 months.</p> <p>--Persons with a job or business in last 4 months who did not work because of layoff or a new job was to begin in 30 days.</p>	<p>--Persons who were jobless in the reference week, available to work, and looking for work at sometime during the last 4 weeks.</p> <p>--Persons who were jobless in the reference week, available to work, but did not work because of layoff or a new job was to begin in 30 days.</p>
Not in labor force	<p>--Persons who neither had a job nor looked for one or was on layoff during last 4 months.</p>	<p>--Persons who neither had a job nor looked for one or was on layoff during reference week.</p>

been absent from their jobs because of illness, vacation, bad weather, labor dispute, and various personal reasons. Unpaid family work is considered employment when it contributes to the operation of a farm or business run by a member of the same household. In the CPS, unpaid family work must have lasted for 15 hours or more during the reference week, but in SIPP there is no hours restriction.

Unemployment. The definitions of unemployment in both surveys are also very similar. In CPS, persons must have been without a job during the reference week and in SIPP they must have been without a job for all or part of the reference period; in addition, they must have been available for work, and taken some specific jobseeking activity. Job-seeking activity in CPS may have occurred anytime in the previous four weeks, while in SIPP it may have occurred any time during the four months. If, in either survey, job seeking occurred when the person was working, working would take precedence and the person would be considered employed.

Two exceptions to the above rule must be noted. The first is the case of the person who has a job but was laid off and the second is the person who is to begin a new wage or salary job within 30 days. Both persons are considered unemployed. In the CPS these persons must have been available for work, but in SIPP no availability test is applied.

Because the CPS is basically a labor force survey, it collects more information about the spell of unemployment than SIPP. For example, CPS gathers information on reasons for unemployment whereas SIPP does not. One can tell from CPS data whether a jobless person has become unemployed because of job loss, such as a layoff; quitting a job to search for another; entering the labor force for the first time; or re-entering the labor force. In SIPP, the only group for whom the reason is known for being unemployed are those persons who report themselves

as having jobs from which they are absent because of layoff. The CPS also asks about the method of job search and how long one has been searching or on layoff. Labor force. The civilian labor force in the CPS is derived by adding the number of persons classified as employed during the reference week to the number who were classified as looking for work or on layoff. The "total" labor force is derived by adding to the civilian labor force an independent estimate of the Armed Forces stationed in the United States.

The labor force in SIPP (which includes members of the Armed Forces living in households but not in installations of the Armed Forces) is referred to as "Persons with some labor force activity." This represents the sum of persons who, during the four month reference period, may have been --

employed during all weeks,
unemployed during all weeks,
employed and unemployed during all weeks,
employed and outside the labor force during all weeks,
unemployed and outside the labor force during all weeks,
and employed, unemployed, and outside the labor force
during all weeks.

In other words, any one with some contact with the labor market in the four month reference period.

Unemployment rate. The unemployment rate from the CPS is one of the most well known statistics in the Nation. It is derived by dividing the number of unemployed persons by the civilian labor force (or total labor force). In SIPP a similar rate, or proportion, could be calculated. Unlike the CPS unemployment rate definition, however, the numerator in the SIPP definition is composed of persons who may have been --

unemployed during all weeks,
employed and unemployed during all weeks,
unemployed and outside the labor force during all weeks,
and employed, unemployed, and outside the labor force
during all weeks.

In other words, the numerator is composed of "Persons with some unemployment."
Dividing the sum of these groups by persons with some labor force activity --
the denominator -- will yield the proportion, or percentage, with some
unemployment. ^{15/}

Not in the labor force. In both the CPS and SIPP, persons who have had no
association with the job market during the reference period (in SIPP, for all
or part of the reference period) are considered outside the labor force. The
CPS further identifies their major activity as in school, keeping house, unable
to work, and so on. This is not done in SIPP.

The CPS inquires in the fourth and eighth rotation groups about previous
work experience, intentions to seek work again, desire for a job, and reasons
for not looking. This makes it possible to estimate the number of "discouraged
workers." Discouraged workers in the CPS are defined as persons who want a job
but are not seeking work currently because: 1) they believe no work is available
in their line of work or area; 2) they could not find any work; 3) they lack the
necessary schooling or training, skills, or experience; 4) employers think they
are too young or old; and 5) they have other personal handicaps in finding a
job, such as transportation problems.

^{15/} The Bureau of Labor Statistics calculates a similar rate from the annual
work experience data collected in the supplement to the March CPS. It is
referred to as "the percent with unemployment."

An effort is made to identify discouraged workers in SIPP also, even though it is difficult to recall a state of mind. For those persons who did not work or look for work in at least part of the four month reference period but said they wanted a job and could have taken one, a question is asked as to why they were not looking. The reasons for not looking are very similar to those in the CPS questionnaire. (See SIPP and CPS questionnaires.)

Hours of work. In the CPS a question is asked about the number of hours some one worked during the reference week at all jobs. This question is asked of all rotation groups and includes workers who have more than one job. In addition, in two of the eight rotation groups a question is asked about the hours "usually" worked at the worker's main job. This information is part of the CPS data collected on workers' earnings.

A similar set of questions is found in SIPP. Everyone who worked is asked about their usual weekly hours on all jobs during the four month period. Subsequent questions inquire about usual weekly hours for the primary job and any others.

Full-time and part-time employment. Full-time employment in both surveys is defined as employment of 35 hours a week or more while part-time employment is anything less than 35 hours. Both surveys seek the reasons for part-time employment, that is, whether it was due to economic reasons or other factors. Economic reasons include slack work, material shortages, repairs to plant or equipment, start or termination of a job during the week, and the inability to find full-time work. "Other" reasons include labor disputes, bad weather, one's own illness, vacation, keeping house, no desire for full-time work, and full time worker during only part of the season. In the SIPP questionnaire the reasons for part-time employment are not as numerous, but it is still possible to identify some economic reasons for part-time employment. (See SIPP questionnaire.)

Uses of SIPP Labor Force Data

SIPP was designed primarily as an income survey and the data from it will be used to address issues related to income security and social welfare programs. With the inclusion of questions on labor force activity, however, this survey has potential for labor force analysis and topics related to it. In addition, because of SIPP's sample design both cross-sectional and longitudinal data can be obtained from the survey providing analysts with more flexibility in their analyses. For example, it is possible to calculate monthly averages of the labor force data from SIPP waves since labor force activity is tracked (week-by-week) over a four month period; on the other hand, by linking all the SIPP waves it is possible to follow the labor force activity of individuals over two and one-half years.

While the CPS will continue to be the primary source of information on the country's labor supply and the current unemployment situation, SIPP labor force data will complement the basic CPS information in many ways. The following is a discussion of some of the applications of SIPP cross-sectional and longitudinal labor force data.

Labor market related economic hardship. For many years economists have tried to measure the economic hardship caused by labor market problems, whether they be demand oriented (unemployment due to insufficient jobs) or supply oriented (low wages because of insufficient skills and education). The economic literature contains many references to subemployment indices, employment and earnings inadequacy indices, and labor market hardship measures of one variety or another.^{16/}

^{16/} For example, see Herman P. Miller, "Subemployment in Poverty Areas of Large U.S. Cities," Monthly Labor Review, October 1973, pp. 10-17; Sar A. Levitan and Robert Taggart, Employment and Earnings Inadequacy: A New Social Indicator, (Baltimore: The Johns Hopkins University Press; 1974); T. Vietorisz, R. Mier, and J. Giblin, "Subemployment: Exclusion and Inadequacy Indexes," Monthly Labor Review, May 1975, pp. 3-12; and Robert Taggart, Hardship--The Welfare Consequences of Labor Market Problems: A Policy Discussion Paper, The W.E. Upjohn Institute for Employment Research, 1982.

The NCEUS in 1979 examined this subject and recommended that the BLS publish an annual report ". . . containing measures of different types of labor market related economic hardship resulting from low wages, unemployment, and insufficient participation in the labor force." ^{17/} Using data from the March CPS, the BLS has produced such reports but they are not as comprehensive as they might be because of data limitations (for example, neither the hourly earnings for part-year workers nor the problems of discouraged workers are discussed.) ^{18/}

SIPP labor force and income data should be able to fill the gap. For example, one cross-sectional table specification might show employment problems incurred by individuals cross-classified by their position in the household income distribution. (See Table 2.) Problems of unemployment, low hourly wages (below the Federal minimum), discouragement, and involuntary part-time employment could be isolated to help in formulating applicable policies. This information, in combination with income information, is available on a current basis only from SIPP.

Labor mobility and turnover. Given the longitudinal feature of SIPP's sample design, not only can the income flows and program participation activities of individuals be monitored for two and one-half years (and periods of shorter duration), but so can their labor force activities. At the time of each SIPP interview, information is obtained on the labor force activity of each household member age 15 and older during the prior four months. Any changes in labor force status during this period are reflected in the data. Stitching together the data collected in each of the eight or nine interviews will provide data

^{17/} See Counting the Labor Force, p. 60.

^{18/} The latest BLS report is Linking Employment Problems to Economic Status, Bulletin 2201, U.S. Department of Labor, Bureau of Labor Statistics, June 1984.

TABLE 2. Persons Age 16 and Over with Employment Problems by Monthly Household Income: Monthly Average, Third Quarter, 1983

Labor force status and employment problems	Total	Persons in households with monthly household incomes of:					
		Under \$ 300	\$ 300 to 599	\$ 600 to 899	\$ 900 to 1,199	\$ 1,200 to 1,599	\$ 1,600 to 1,999

Total persons 16 and older

Persons with some labor force activity or interest^{1/}

Persons with no employment problems

Persons with some employment problems

Unemployment

Discouragement^{2/}

Low wages^{3/}

Economic part-time employment^{4/}

Unemployment and discouragement

Unemployment and low wages

Economic part-time and low wages

Economic part-time and unemployment

All other combinations of problems

Percent of persons with some labor force activity or interest who have employment problems

Persons with no labor force activity or interest

- ^{1/} Persons with labor force interest are those who did not work during the month and were classified as discouraged workers. See footnote 2 below.
- ^{2/} Discouraged workers are persons who wanted a job and were available to work but did not look for work because 1) they thought no work was available in their line or area, 2) they could not find any work, 3) they lacked necessary schooling, training, etc., 4) employers thought they were too young or too old and, 5) any other personal handicaps.
- ^{3/} Low wages are defined as an hourly wage at or below the Federal government's minimum wage. For workers not paid by the hour, the average hourly wage is computed by dividing monthly earnings on the main job by the product of the usual weekly hours times the number of weeks in the month.
- ^{4/} Persons with economic part-time employment are those who worked less than 35 hours in some or all weeks during the month because they could not find a full-time job or because of slack work or material shortages.

users with a profile of labor market activity for a two and one-half year period.

One change in labor force status that labor economists have been interested recently is the one which occurs after a spell of unemployment. Some have argued that many outcomes of spells of unemployment are withdrawals from the labor force and not reemployment. For example, two economists using CPS gross flow data, estimated that 45 to 50 percent of all unemployment spells end by labor force withdrawal.^{19/} Other economists have argued that the relative shortness of the average unemployment duration shows that persons can quite easily find their usual type of employment in a short period of time.^{20/} With SIPP labor force data it will be possible to identify job terminations, observe spells of unemployment, and determine not only their durations, but their outcomes.

SIPP labor force data should also be useful in calculating rates of job separation and accession. The measurement of the amount of job separation and accession is an important element in understanding our basic employment and unemployment statistics. Since the discontinuance of the BLS's labor turnover series, researchers have been hard pressed to find other data sources which would shed light on the dynamics of the labor market.^{21/} While it will not be possible

^{19/} See Kim Clark and Lawrence Summers, "Labor Market Dynamics and Unemployment: A Reconsideration," Brookings Papers on Economic Activity, No. 1, 1979, pp. 13-72.

^{20/} See Martin Feldstein, "The Importance of Temporary Layoffs: An Empirical Analysis," Brookings Papers on Economic Activity, No. 3, 1975, pp. 725-744.

^{21/} For a statement of the need for labor turnover data, see Robert E. Hall and David M. Lilien, "The Measurement and Significance of Labor Turnover," in Counting the Labor Force, Appendix Vol. 1 (Concepts and Data Needs), National Commission on Employment and Unemployment Statistics (Washington D.C., 1979), pp. 577-600. For an example of separation data created from the CPS see S. Haber, E. Lamas, and G. Green, "A New Method for Estimating Job Separations by Sex and Race," Monthly Labor Review, June 1983, pp. 20-27, and Allan Eck, "New Occupational Data Improve Replacement Estimates," Monthly Labor Review, March 1984, pp. 3-10.

from the SIPP labor force data to identify the precise nature of the separation (layoff, quit, discharge) or accession (new hire, recall), aggregate separation and accession rates could be calculated. These rates could be monitored over the business cycle.

Summary

SIPP is principally an income survey, but it contains questions on labor force activity as well. SIPP labor force data will supplement the labor force information from the CPS, the Federal government's official source of labor force statistics. Like the CPS, SIPP uses an activity concept for sorting the Nation's population into those persons involved in the job market from those who are not. A major difference between the two surveys is the length of the reference periods; in the CPS it is one week and in SIPP it is four months. The different length of time for which labor market activities are surveyed will be an important factor in SIPP and CPS labor force comparisons. Nevertheless, while the CPS will continue to tell us how many persons are employed and unemployed, SIPP will be able to tell us how well the labor market is providing for these workers and their households.

Matching Economic Data to the Survey of Income and Program Participation:
A Pilot Study

Sheldon E. Haber*
Paul M. Ryscavage**
Douglas K. Sater**
Victor M. Valdisera**

Preliminary Report

Paper presented at the American Statistical Association meetings,
Philadelphia, PA., August 1984

* The George Washington University.
** Bureau of the Census.

Table of Contents

	<u>Page</u>
Introduction	1
SIPP and the Economic Data	4
Survey of Income and Program Participation	5
The Standard Statistical Establishment List	8
The Longitudinal Economic Data File	11
Enterprise Statistics	14
Some Applications of Micro-Demographic and Economic Data	16
The Relationship Between Wage Rates and Capital	16
Labor Mobility	22
Low Wage Workers and Low Wage Firms	25
Measuring the Effects of Minimum Wage Legislation	28
Structural Unemployment	34
Identifying High Tech Workers and High Tech Firms	35
Implications of the Transition from a Goods to a Service Economy	38
The Pilot Study	41
Methodological Problems in Matching Demographic and Economic Data	42
Procedures for Identifying an Individual's Place of Work	43
Estimating Missing Economic Data for Small Establishments and Companies	45
Footnotes	49
References	50

Preliminary Report

Matching Economic Data to the Survey of Income and Program Participation:
A Pilot Study

Sheldon E. Haber*
Paul M. Ryscavage**
Douglas K. Sater**
Victor M. Valdisera**

The new Survey of Income and Program Participation (SIPP) will undoubtedly become a major source of data on a wide variety of aspects of the well-being of our nation's households, families, and individuals. SIPP is designed to collect information about cash and noncash sources of income, taxes, and assets and liabilities from which improved estimates of income, poverty, and wealth can be derived. While a principal focus of the survey is on income and program participation with the intent of assessing policy issues such as the effects of proposed changes in program eligibility rules on benefits, it will address a much wider range of policy questions and yield data for analytical studies in a variety of areas of economic inquiry. The very richness of SIPP suggests the desirability of augmenting it with micro-level establishment and enterprise data from the economic censuses and other data files maintained by the Bureau of the Census, since the marginal cost of merging these data with SIPP is relatively small and the potential gain in knowledge is very large.

One area where the payoff relative to cost of enhancing SIPP is sure to be substantial and significant is that pertaining to the behavior of labor markets. A primary source of data for verifying established propositions relating to labor market phenomena and exploring new ones has been the Current Population Survey (CPS). What distinguishes SIPP from the CPS is that the latter is cross-sectional while the former is longitudinal. Other well known longitudinal

* The George Washington University.

** Bureau of the Census.

data sets exist, e.g., the National Longitudinal Survey maintained by Ohio University and the Social Security Administration Longitudinal Employee-Employer Data (LEED) file; however, these do not have the same breadth of coverage as SIPP. Each of these data sets provides information almost exclusively about workers but very little or no information about the places in which they work. By bringing together information about workers and their place of employment in a single data set, denoted below as the SIPP-Economic Data (SIPP-ED) file, a major gap can be closed in our knowledge of how labor markets function. Additionally, the file can be expected to add new insights into firm production functions.

A list of some of the areas in which a SIPP-ED file can yield new insights includes the following topics:

- The relationship between capital and wage rates

- Labor mobility

- Low wage workers and low wage firms

- Measuring the effects of minimum wage legislation

- Structural unemployment

- Identifying high tech workers and high tech firms

- Implications of the transition from a goods to a service economy

- Unions and the labor market

- The substitutibility of capital and labor

- Productivity analysis

The merging of demographic and economic data will enable investigators to obtain improved estimates of the impact of economic and institutional forces which have been intensively studied but still are only partially understood. It will also enable investigators to examine aspects of labor market outcomes and production processes that have, heretofore, been difficult to study.

Besides the substantive knowledge to be gained by merging SIPP and economic data, there are externalities associated with merging these data sets. First, it will be possible to verify the accuracy of the size of firm estimates given by respondents in survey data. Unpublished comparisons made by one of the authors of respondent estimates of firm size from the May 1979 CPS with control totals from the economic censuses suggest that there is underreporting of firm size. The economic files also permit one to more accurately identify the industry in which a worker is employed since census industry codes do not match neatly with SIC codes. As Mellow and Sider (1982) have shown, industry designations by survey respondents and employers differ approximately 8 and 15 percent of the time at the 1- and 2-digit SIC levels, respectively. When economists proxy monopoly power using industry concentration ratios in conjunction with CPS data, they must average the ratios for 2- or 3-digit SICs to match the census industry classification. By merging demographic and economic data, the more accurate 4-digit SIC industry concentration ratios in the source data can be utilized.

An additional, indirect benefit of linking demographic and economic data stems from the fact that the former represent a representative sample of the working population. Matching on work place will yield a stratified sample of firms where the probability of selection is inversely proportional to firm size. By weighting the number of firms in each size group, estimates for the entire population of firms can be derived. The sample of employers would be contained in a single data set--versus the diversity of data sets in which the economic data are now found -- with the same format across employers. These advantages plus the manageable size of the sample should provide valuable insights into the structure of production within and across sectors of the economy at a point of time and over time.

As is the case in designing and improving any data base, it is essential to have, at the outset, a clear idea of the study areas and issues to which it may be applied. The primary objective of this paper is to provide such a focus. This is done in section 2 where applications of the proposed SIPP-ED file are surveyed. All but the last three topics noted above are discussed; the latter will be incorporated in a final version of this report. Before examining the applications of a SIPP-ED file, attention is first given in section 1 to the kinds of information in SIPP and the major economic data files maintained by the Bureau of the Census, since it is the information in these files which would form the data set to be used in implementing the studies described below. In section 3, additional aspects of the pilot study are outlined. Two methodological problems which need to be addressed in developing a SIPP-ED file are examined in section 4.

1. SIPP and the Economic Data Files

In merging demographic and economic data, it is necessary to know the information contained in the various files to be linked and how each file is constructed. In this section we briefly describe four data sets which might be incorporated into a SIPP-ED file. As mentioned, the demographic data are contained in SIPP. Economic data are found in the Standard Statistical Establishment List (SSEL), the Longitudinal Establishment Data (LED) file, and the enterprise statistics (ES). The SSEL covers all establishments and companies with employees and yields current information on employment and payroll. The LED, as its name implies, contains longitudinal data but is restricted to manufacturing establishments. The ES, on the other hand, covers companies in the construction, mineral, manufacturing, wholesale trade, and retail trade industries, and most service industries.¹ Each of these data files is discussed below.

A. The Survey of Income and Program Participation

SIPP is an ongoing series of national panels from which income and program participation information will be derived. Currently, the main source of such data is the March income supplement of the CPS, but because the principal purpose of the CPS is to obtain information on employment and unemployment, a more refined instrument is needed. SIPP is designed to obtain improved reporting of income and participation in major income security programs, as well as to expand on information needed to fully analyze program participation and eligibility.²

In each SIPP panel approximately 20,000 households will be interviewed over a two and one-half year period. Since panels will overlap, cross-sectional estimates can be obtained for a combined sample of about 40,000 households. To reduce the interviewing workload, each panel is divided into four rotation groups of about equal size. One rotation group is interviewed during the first two weeks of each month. One cycle or wave of interviewing of the four rotations requires four months; thus, each household is interviewed three times a year. The reference period for an interview is the four month period preceding the interview month, e.g., the reference period for the interview month October is June through September.³

SIPP consists of four parts. The first is a control card containing such information as age, sex, race, ethnic origin, marital status, educational level, veteran status, place of residence, and names of employers. The second contains a set of core questions covering labor force participation and amounts and types of income received during the reference period. In addition to wage and salary income and income from self-employment, questions are asked about cash transfer payments from governmental programs such as social security, disability,

unemployment benefit, and welfare programs. Information on the receipt of non-cash benefits from Food Stamps, Medicare, and Medicaid is also obtained. Other core income questions relate to private transfers, e.g., pensions from employers, alimony, and child support.⁴

The last two parts are fixed topical modules and variable topical modules. The topics covered in these modules do not require repeated measurement during the year and the reference period may be longer than the four month period used for the core questions. The fixed topical modules may be repeated during a panel and over successive panels. For example, the wealth (assets and liabilities) module will be administered twice in each panel, in waves one year apart. A "round-up" module is administered at the end of the first and second years of interviewing to obtain annual estimates of wage and salary income, self-employment income, and property income; estimates of taxes and employer provided benefits are also obtained in this module. Other fixed topical modules provide information in the following areas: work history, education history, and health and disability (denoted below as the work history module); educational enrollment; and marital history, fertility, and migration. The variable topical modules include supplemental questions designed by or for other federal agencies. Although these questions may be repeated from one panel to another, they often are of such special interest that they may be asked only one time over a span of years. Included among the variable topical modules are questions relating to reasons for not working (in which information on a worker's reservation wage will be collected); pension and retirement issues; work related expenses; child care arrangements and financing; health care utilization and financing; housing conditions and costs; energy usage; and other topics.⁵

For the purpose of indicating how economic data can materially augment SIPP, it is sufficient to note that SIPP contains retrospective and prospective labor market information which goes much beyond the scope of the CPS and other longitudinal data sets. Besides the information that is normally found in the CPS with respect to labor force attachment and employment status,⁶ the SIPP core questionnaire contains information on the usual number of hours worked per week by an individual and his or her rate of pay for each employer. Additional information about each individual is found in the fixed topical modules. In the work history module, for example, questions relating to the following areas are asked:

Education history

- Program of studies taken in high school
- Courses taken in high school
- Highest degree attained beyond a high school diploma
- Field of study of highest degree
- Year highest degree received

Work history

Training

- Source of latest training (e.g., apprenticeship program, training program at work, military training)
- Date and length of training program
- Who paid for the training (e.g., self or family, employer, government)

Prior work experience

- Year in which first worked at a job lasting six consecutive months or more
- Number of years worked six or more months during the year
- Typical work status during years worked (full-time, part-time)
- Number of times, duration, and reason for not working six or more consecutive months

Last job (for persons working 10 or fewer years at current job)

- Year started and ended job
- Usual hours worked
- Rate of pay at end of job
- Reason for leaving last job (e.g., layoff, discharge, other reason)
- Time between last and current job

Current job

- Size of establishment and firm
- Single or multi-establishment firm
- Union status and coverage by a union contract
- Rate of pay at start of current job
- Years worked for current employer
- Years worked in current occupation for current employer

Health and disability

- State of health and if disabled, length of time disabled

As indicated above, although SIPP provides extensive historical data about an individual's work experience, little information is available about the firms in which that experience is gained. The economic data described below would remedy this shortcoming.

B. The Standard Statistical Establishment List

The SSEL is a complete directory of establishments in single and multi-establishment enterprises⁷ with one or more employees, irrespective of industry. The SSEL links parent companies, subsidiaries, and their establishments. It contains information on approximately 4.7 million enterprises and 5.7 million establishments.

The SSEL is comprised of three files. The first is the Single Unit (EC-EI) file which contains information for businesses with paid employees, which, therefore, must file for an employer identification number (EIN) with the Internal Revenue Service. Such businesses, represented by their EIN, may consist of a single establishment company; an establishment or subsidiary (with one or more establishments) of a multi-establishment firm which itself can be part of a larger enterprise; or an entire multi-establishment parent company. All establishments

belonging to a multi-establishment parent company and the establishments of its subsidiaries are listed under the parent company's identification number, denoted as an alpha number,⁸ in a second file, the Multiple Unit (EC-MU) file. The third file, known as the Master Mailing Address file, contains the mailing addresses of each enterprise. The SSEL also contains the address of the physical location of each establishment.⁹ The addresses for single establishment enterprises are found in the EC-EI file; the addresses for establishments in multi-establishment enterprises are found in the EC-MU file.

The importance of the SSEL is that it is a current file containing a complete list of establishments and companies with paid employees. This contrasts with administrative records maintained by the Social Security Administration as reflected, for example, in the LEED file in which the unit of observation is a business defined in terms of an EIN. In this case it is not clear whether the EIN refers to a single establishment or a group of establishments. Moreover, if an EIN changes, as may occur when a business changes ownership, it will appear that it has closed down. Analogously, employees of such businesses will appear to have changed jobs. In the SSEL, each establishment is identified by a unique census file number (CFN)¹⁰ which remains the same even when there is a change in its EIN or the EIN of the subsidiary or enterprise to which it belongs. For longitudinal studies in which workers and/or establishments are tracked over time, it is essential that identification numbers remain invariant. This function is fulfilled by the Social Security Number for workers and by the CFN for establishments.

While the SSEL contains a narrow range of economic data, these data impart valuable information. The establishment data contained in the SSEL are as follows:

Identification

Name of establishment (and company)

SIC 4-digit industry

Location identifiers (e.g., address of physical location of establishment)

Number of employees, 1st quarter

Payroll, 1st quarter and annual

Sales and receipts¹¹

Legal form of organization (e.g., sole proprietorship, partnership, corporation)

Date EIN was entered into the SSEL

Reason for EIN being issued (e.g., started a new business, change in ownership)

Operational status (e.g., active, inactive)

The address of the physical location of an establishment is useful for the merging of demographic and economic data, since it is a primary link in identifying an individual's place of work. Identification of the establishment in which a person works enables one to determine his or her industry at the 4-digit SIC level. The date an EIN was entered in the SSEL and the reason for it being issued provide information about the age of an establishment. The information on employment yields a more accurate estimate of employer size than that which can be obtained from respondent's estimates in survey data. The employment and payroll figures also yield an estimate of average annual earnings,¹² thereby indicating whether an employer is a low or high wage employer. And the sales and employment figures provide a proxy measure of productivity. The operational status information can be utilized to identify establishments which have become inactive. It should be noted that the SSEL contains longitudinal information. Currently, establishment and company data are carried for two years in the SSEL.¹³

C. The Longitudinal Establishment Data File

The LED is a longitudinal micro-data base containing data at the establishment level from the Annual Survey of Manufactures (ASM) and the Census of Manufactures (CM). The data begin in 1972 and currently extend through 1981; it is anticipated that data from the 1982 CM and 1983 ASM will be incorporated into the file by the end of 1984. The LED was developed jointly by the Bureau of the Census and Yale University under the direction of Richard and Nancy Ruggles.

Every year economic data are collected in the ASM from a sample of 55,000 establishments. Included in the ASM sample are all establishments with 250 or more employees.¹⁴ Thus, continuous longitudinal data are available for all large manufacturing plants. Depending on industry, a sample of establishments with between 5 and 20 but less than 250 workers is included in the ASM; after five years these establishments are replaced by a new sample of small plants. For the very smallest manufacturing plants, i.e., those with less than five to 20 workers, again depending on industry, economic data are obtained from administrative records of other government agencies, e.g., the Social Security Administration, rather than from a questionnaire.

Every fifth year a Census of Manufactures is taken. The establishments in the ASM are the core of the CM. All establishments with 5 to 20 or more employees which are not in the ASM are sent a CM questionnaire.¹⁵ As in the ASM, the economic data for the very smallest establishments are derived from administrative records. The year-to-year linkage of data in the LED is poorest for the very smallest plants and improves as plant size increases. Likewise, the breadth of data improves as plant size increases. For the very smallest plants, the only economic data available in the LED are those contained in the SSEL. A much wider range

of data are available for plants responding to the ASM and CM questionnaires. The latter questionnaire contains the following information:

Identification

CFN number
SIC 4-digit industry
Location identifiers (e.g., state, SMSA, county)

Legal form of organization

Number of employees

Production workers (average as of the pay periods including the 12th of March, May, August, and November)
All other employees (pay period including March 12th)

Payroll

For production workers and for all other employees (annual)
All employees (1st quarter)

Hours worked by production workers (for each quarter)

Cost of materials and services used (including cost of fuels consumed for heat and power and of purchased electricity, cost of contract work done by others)

Inventories, beginning and end-of-year

Capital expenditures (for new structures and for new machinery, used buildings and machinery)

Operational status (e.g., active, temporarily inactive, ceased operation)

Total value of shipments

The above information as well as that shown below is collected from all large establishments and a sample of small establishments responding to the ASM questionnaire:

First year of operation (1975 and 1981 only)

Supplementary labor costs

Legally required (includes social security tax, unemployment tax, workmen's compensation, and state disability tax)

Voluntary programs (includes life and medical insurance premiums, payments into pension and welfare plans, union negotiated benefits, and payments into stock purchase plans).

Gross value of depreciable assets, beginning and end-of-year (for structures and for machinery)

Gross value of retired depreciable assets (includes assets sold, retired, scrapped, destroyed, etc.)

Depreciation charges (for structures and for machinery)

Rental payments (for structures and for machinery)

Value of shipments to other plants of the same company

As seen from the figures below, 63.4 percent of the value of shipments in manufacturing originates in plants with 250 or more employees. These large plants, comprising 4.0 percent of all manufacturing plants, employed 56.6 percent of the work force in 1977.

Percent Distribution of Manufacturing Establishments with Paid Employees by Size Class
(1977 Census of Manufactures)

<u>Size of Establishment (Employees)</u>	<u>Number of Establishments</u>	<u>Number of Employees</u>	<u>Value of Shipments</u>
Less than 50	82.2	15.3	11.8
50 - 99	7.6	10.1	8.6
150 - 249	6.1	18.0	16.2
250 - 499	2.4	15.6	15.0
500 - 999	1.0	13.5	14.8
1000 or more	0.6	27.5	33.6
Percent	100.0	100.0	100.0

As indicated, the LED provides a much broader range of information about establishments than the SSEL. For all manufacturing establishments, it is possible to derive an accurate estimate of the average wage of production workers. For all but the very smallest establishments, value added¹⁶ per production worker, which is a measure of labor productivity, can be calculated. For the larger establishments, information is available on depreciable assets and rented machinery

so that capital/labor ratios can be computed. Also, a better measure of labor compensation, including fringe benefits, can be obtained. The degree to which a plant is vertically integrated with other plants can also be inferred.¹⁷

D. Enterprise Statistics

Like the Census of Manufactures, the enterprise statistics (ES) are collected every five years. The latest ES data are for 1982. These data cover enterprises whose primary activity is in an in-scope industry.¹⁸ For each enterprise, the data are consolidated over all operating units. The information contained in the ES is similar to that in the Census of Manufactures. The following information, derived from the economic censuses, is available for all enterprises:

Identification

- Alpha number
- SIC 4-digit industry
- Location identifiers

Legal form of organization

Single industry or multi-industry¹⁹

Number of owned establishments

Number of employees²⁰

Payroll, annual

Total sales and operating receipts (excluding the value of intra-company transfers among own establishments)

Value added (only for a company's establishments in the mineral, construction, and manufacturing industries)

Inventories, beginning and end-of-year

Capital expenditures (for new structures and for new machinery, used buildings and machinery)²¹

- New computers and peripheral data processing equipment
- Other expenditures for new machinery

The above information as well as that shown below is collected for large enterprises, defined as companies with 500 or more employees, using a separate enterprise questionnaire:

Cost of purchased advertising

Supplemental labor costs

Legally required
Voluntary programs

Gross value of depreciable assets, beginning and end-of-year (for structures and machinery)

Other domestic assets for which depreciation or amortization reserves are not maintained

Net value of depreciable assets

Net foreign assets

Total depreciation charges

Rental payments (for structures and for machinery)

Assets acquired through capital leases in 1982

The distribution of companies by size class is given below. As indicated, companies with 500 or more employees comprise 0.3 percent of all companies with paid employees. These large companies accounted for 47.5 percent of all employees and 48.3 percent of sales and receipts in 1977.

Percent Distribution of Enterprises with Paid Employees by Size Class
(1977 Enterprise Statistics)

<u>Size of Enterprise (Employees)</u>	<u>Number of Enterprises</u>	<u>Number of Employees</u>	<u>Sales and Receipts</u>
None ^a	10.3	-	0.8
1-49	86.5	32.8	31.7
50-249	2.7	15.0	14.8
250-499	0.2	4.8	4.5
500-999	0.2	4.2	4.4
1000 and over	0.1	43.3	43.9
Percent	100.0	100.0	100.0

^{a/} Companies which reported annual payroll, but did not report any employees on their payroll for specified pay periods in 1977 (see footnote 20).

Two items of interest which are contained in the 1982 ES but not in the LED are capital expenditures for computers, the cost of purchased advertising, and net foreign assets. The first of these, capital expenditures for computers, is of interest because it indicates use of a technology which underlies a number of new industries, often described as high tech industries. The second, the cost of purchased advertising, is a partial measure of the degree to which a product market is competitive.

With this perspective of what is available in the economic censuses for individual establishments and companies, we turn now to a more detailed discussion of the applications of a SIPP-ED file. The reader should bear in mind the limitations of the data, in particular, the lack of asset data for smaller establishments and companies. While such data for are not universally available, this does not preclude the possibility that the data base can be augmented in the future, e.g., by developing an analytical model for imputing assets.

2. Some Applications of Micro-Demographic and Economic Data

In this section, a more detailed discussion of the applications of a SIPP-ED file is provided. The main objective is to show how the uses of such a data set cut across labor market and production theory. In pursuit of this objective, we focus on a number of issues and hypotheses, some of which have been examined in the literature using data sets whose informational content is not as rich as the one to be developed in the pilot study; others have yet to be explored because the necessary data are unavailable.

A. The Relationship between Wage Rates and Capital

Perhaps the most immediate application of a SIPP-ED file data is in the area of wage determination. Despite the voluminous number of analyses in which investigators

have sought to explain wage rates by sex, race, union status, and region, our understanding of the wage determination process still remains incomplete. One area which warrants further investigation is the relationship between wage rates and capital. While the productivity of labor is strongly related to the amount of capital with which it is combined, we have been able to identify only a few wage rate studies that incorporate variables relating to capital use. One reason for this may be that such data are difficult to obtain, but this cannot be the only reason since asset data are published in the CM and investigators have utilized other information, e.g., industry concentration ratios, which is equally difficult to access. Another reason is that economic theory suggests that competitive wage rates are independent of the amount of capital utilized by a firm. Important exceptions to this proposition, however, are worth noting. To gauge the significance of these exceptions, economic data are required.

In theory, variations in capital among firms have no impact on wage rates provided that there are no market imperfections. In a competitive labor market, any firm can hire as many units of a given quality of labor as it requires at the prevailing market wage. The ability of a firm to do this does not depend on the amount of capital it utilizes in its production process (or any other attribute of the firm) as long as the amount of labor it hires is not sufficiently large relative to the amount available to influence the wage rate. Although some firms have more resources to purchase labor because they are efficient, e.g., because they are better able to incorporate capital into their production process, the fact that they are efficient means that their output will be larger than that of inefficient firms; as a consequence, they will tend to use those resources to purchase higher quality labor,²² again at the prevailing market wage for such labor, rather than pay more than the market wage for lower quality labor.

Several premises underlie the competitive model which may not be met in practice, e.g., that labor quality can be precisely defined and accurately measured so that firms are able to determine with exactness which quality of labor a particular applicant belongs to. The premise that a firm can accurately determine worker quality is cast into doubt by the large sums of money that are spent in screening applicants for employment.²³ Since more efficient firms can still survive even though they may overpay some workers whose quality is overestimated, less efficient ones may only be able to survive if they are successful in paying workers only what their quality warrants. Thus, a positive relationship between wage rates and firm productivity is plausible, even in competitive labor markets.

It is also clear that no matter how precisely an occupation is defined, workers within that occupation will differ in terms of their work effort, attentiveness, attendance record, and similar attributes not measured by survey instruments but for which firms are willing to pay a premium. Capital intensive firms which tend to hire high quality workers, some of whose characteristics are difficult to observe, should, therefore, exhibit higher wages than labor intensive firms, again, even when labor market competition is present. One implication of this proposition, which is supported by findings reported by Brogan and Erickson (1975), is that the (positive) relationship between the composite wage rate in an occupation and the amount of capital utilized by a firm should be strongest among occupations directly associated with capital, i.e., the skilled occupations, and weakest for those occupations where this association is weakest, i.e., the unskilled and clerical occupations.²⁴

Market imperfections are another reason why economic data are relevant to the wage determination process. Imperfections in the product market result in firms securing higher than normal profits. Imperfections in the labor market

result in unions attempting to share in firm profits. Both phenomena are related to firm size: large firms are able to concentrate market power and pass on cost increases; they are also easier to unionize.

The relationship between wage rates and firm size has been documented in a number of studies.²⁵ Using industry data, Masters (1969) found that one-fifth of the variation in the hourly earnings of production workers in manufacturing was explained by plant size. The remaining variables, including the extent of unionization and concentration in an industry, raised the percentage explained by only another one-fifth. Mellow (1982b), using the May 1979 CPS and pension supplement containing information on firm size, has found that compared

"to the excluded plant and firm size category (less than 25 workers) ... the combined wage premium for an average worker in the largest plant ... and company size category ... is 23 percent." (p. 497)

This estimate of the impact of size of business on the wage rate, 23 percent, is as high as the much more heralded union effect.

Size of firm captures a host of relationships that affect wage rates. Indeed, that is the problem with using a variable which captures much but reveals little by itself. To unlock the puzzle between firm size and wage rates requires information about the characteristics of individual firms. The most important characteristic may be the amount of capital a firm utilizes. When product markets are not competitive because of the presence of large firms, labor can capture some of the gains from working with more (versus less) capital through unionization or the threat of unionization. To the extent that this is so, workers of the same quality will receive a higher wage simply because they have relatively large

amounts of capital to work with. In this case, the wage rate and capital/labor ratio will be positively related. Moreover, where the capital/labor ratio is positively related to other variables which themselves impact on the wage rate, the wage elasticity with respect to these other variables will decline. In particular, if there are complementarities in production between higher quality labor, e.g., better educated labor, and capital, the rate of return to education will be lower when the capital/labor ratio is included in the wage model than when it is omitted. The same will be true with respect to the wage gains of labor unions. If a labor union's ability to raise wages is enhanced in capital intensive firms, the union-nonunion differential will be overestimated if variations in the capital/labor ratio among firms are not taken into account.²⁶

Besides the direct effect of the capital/labor ratio on the wage rate, interaction effects with other variables are to be expected. For example, an interaction effect may be expected between on-the-job training (OJT) and the capital/labor ratio. It is well known that the age-earnings profile with respect to OJT depends on the extent to which it is specific to a firm or general to many firms. Since specific training tends to be paid for by the firm, the age-earnings profile is flatter for this kind of training. Of practical import is the problem of determining whether a given kind of OJT is specific or general. The capital/labor ratio provides a way of proxying the specificity of OJT. It is likely that in establishments where the capital/labor ratio is high, the capital is sophisticated and performs much of the work. The role of the employee is to insure that the equipment is maintained in good order and its capabilities are fully utilized, but to succeed in these tasks specific training is often required. To the extent that this is so, while the wage rate should vary positively with OJT, the interaction effect (on the wage rate) of OJT and the capital/labor ratio is

likely to be negative. Additionally, the rate of return to education depends on whether it is combined with OJT that is specific or general. If it is combined with specific OJT, the age-earnings profile will be flatter and the rate of return less than if it is combined with OJT that is general. But as indicated, OJT should be more specific the higher a firm's capital/labor ratio.

Of significance for the discussion at hand, not all large firms have high capital/labor ratios. For example, large firms in retail trade have smaller capital/labor ratios than those in manufacturing. Moreover, the relationship between the capital/labor ratio and earnings may differ across industries, everything else held constant. It seems reasonable to assume that the kind of capital a firm invests in will also affect worker productivity. For example, one might expect wage rates to depend on whether the most recently acquired assets are new or used, the rate at which assets are retired, and the proportion of capital expenditures invested in new computers. For these reasons, information about assets, and capital expenditures, should lead to a reduction in the percentage of the variation in wages which is left unexplained in wage rate models.

In addition to differences in the mix of capital and labor, firms also differ in their ability to pay high wages. Ability to pay has been associated with the degree to which employment in an industry is concentrated in a small number of firms.²⁷ Ability to pay may also be evidenced by the age of a firm. All else being the same, young firms attempting to gain a foothold in an industry may not be able to pay as high a wage to attract labor as older firms.²⁸ Wage

rates may also depend on changes in the level of employment; they are likely to be lower in firms that have experienced substantial contractions in their work force than in firms that are undergoing vigorous growth in employment.

Still another advantage to be realized from the economic data is the availability of information pertaining to supplementary benefits for voluntary programs. These fringe benefits contribute substantially to total compensation²⁹ and should be taken into account in explaining labor market outcomes. Antos (1983) has found for example, that when nonwage compensation is ignored, the union impact on employee income is seriously understated. As indicated, fringe benefit information, as well as other economic data mentioned in this section, would be contained in a SIPP-ED file.

B. Labor Mobility

The literature on labor mobility has been primarily concerned with the factors that influence workers to change employers and the effects of such mobility on individuals, employers, and the economy.

As Mincer and Jovanovic (1981) have shown, the quit rate decreases with years of experience in the labor force and with the length of time individuals work for their employer. Since specific training tends to be paid for by the employer firms have an incentive to retain workers who receive such training. One way of doing this is by paying higher wages and/or by offering more extensive fringe benefits, e.g., pension benefits. Workers are also reluctant to leave an employer because specific training is not easily transferable among firms.

Implicit in our discussion of the determinants of the wage rate is the hypothesis that the higher the capital/labor ratio, the stronger is the relationship between job tenure which is observable and specific training which is not observable.

Thus, holding job tenure and other variables constant, workers in firms with a high capital/labor ratio are likely to have a lower quit rate than similar workers in firms with a low capital/labor ratio.

Quit rates also have been found to be lower in those industries in which fringe benefits are a large proportion of total compensation (Pencavel, 1970). Since the economic data provide information by firm on voluntary supplementary labor costs, improved estimates of the elasticity of the quit rate with respect to these costs (which represents benefits to employees) will be possible.

Labor mobility is also central to the question of dual labor markets. Dual labor market theory suggests that low income is not only related to the characteristics of individuals, in particular, to inadequate or inefficient investment in human capital, but also to the characteristics of the jobs which they hold. In this view there are two kinds of labor markets. In the primary sector, jobs are characterized by high and increasing pay, job security, and on-the-job training. The secondary labor market, in contrast, consists of low wage jobs with few opportunities for advancement, high turnover, and little skill development.³⁰ Individuals who are trapped in the secondary labor market are believed to earn less than their counterparts with similar characteristics in the primary sector. This lack of labor mobility between low and high wage employers, which is the essential feature of dual labor market theory, can be tested using SIPP and economic data from the SSEL.

Although respondent estimates of establishment and firm size are found in some data sets, such information is generally available only for a particular moment in time. This is also true of SIPP. However, by combining SIPP with SSEL data, firm size as well as other employer characteristics can be developed for all of an individual's employers.

One area where such information is crucial relates to human capital transfers between firms of different size. Schiller (1982) has suggested that by their very nature small firms expose workers to a variety of job skills and training under intensive supervision. But since small firms pay less than large ones for the same skills, workers who receive training in the former move to the latter. Schiller estimates that "small firms are losing over 30,000 newly-trained workers (net) each year to larger firms" (p. 68). According to Schiller, although individual workers and large firms benefit from this transfer of human capital, small firms lose with the resultant outcome that they may provide less training than otherwise.

Schiller's study is based on data from the Social Security Administration LEED file and, thus, suffers from the limitation that the EIN provides an imperfect measure of labor mobility and firm size. Moreover, training is measured indirectly by relative increases in the wage rate. The LEED file, however, does not indicate the hours worked by an individual; hence, individuals who switch from part-time to full-time work will experience large relative gains in earnings, and the amount of training they receive will appear to be larger than is actually the case. The overstatement of training will be most pronounced for small firms, since part-time work is most prevalent for this class of employers. Besides information on part-time employment, a SIPP-ED file would contain information on education, occupation, union status, job tenure, employment experience, marital status and other variables that are missing in the LEED file and should be taken into account when explaining human capital formation. Failure to include these variables, which are correlated with firm size, can lead to misestimation of the amount of training provided by small and large businesses.

With the availability of micro-worker data, investigators have begun to look at the return to job mobility, in particular, whether job changers experience more rapid wage growth than job stayers. For men, it appears that although job separations lead to short-run gains in wage rates, in the long-run the largest gains are registered by those who stay with an employer (Borjas, 1981).

While this may be true in general, it may not be true with respect to important classes of employers, e.g., low wage employers. Whether this is so is clearly of some interest in furthering our understanding of how labor markets function.

C. Low Wage Workers and Low Wage Firms

The relationship between poverty and low wage rates is self-evident. In 1978, the average poverty threshold for a family of four was \$6,662.³¹ In the same year, almost 40 percent of the wage and salary workers in families with incomes of less than \$6,000 earned the minimum wage or less.³² The characteristics of these low wage earners are the same as those of persons living in poverty, i.e., a relatively high proportion are young or old, black, and female.

While survey data such as the CPS provide insights into the characteristics of low wage workers, they provide no information about low wage firms. Given that a firm pays low wages, the price of labor relative to capital will tend also to be low. All else being the same, such firms will be labor intensive and, hence, tend to be smaller than high wage, capital intensive firms.

And because recruitment and hiring costs relative to the level of wages will tend to be high, such firms will also advertise less for labor and employ fewer screening devices to weed out unsuitable workers; thus, their work force will be of lesser quality than their high wage counterparts. Marginal workers, on the other hand, e.g., younger workers and those who are less

educated, will be attracted to low wage firms because their marginal product is less than that required to gain employment in high wage firms. More generally, workers with given characteristics and tastes sort themselves among firms with similar requirements for labor. The outcome of this process is an equilibrium relationship between compensation, on the one hand, and worker and firm characteristics, on the other hand, as determined by successful job matches.

Corresponding to the greater prevalence of low quality workers in low wage firms, one might expect that in these firms (vis-a-vis high wage firms) a higher proportion of capital expenditures is for used rather than new machinery and equipment; likewise, the proportion of depreciable assets retired each year is likely to be smaller in such firms. Furthermore, given that labor is of lesser quality and capital is of an older vintage, it would not be surprising if value added per worker were relatively low in low wage firms.

Other characteristics are more easily seen by focusing on high wage firms. To the extent that high wage firms are capital intensive, their need for trained workers is likely to be greater than that of low wage firms. Capital intensive-ness suggests greater use of resources to monitor output; hence, a higher proportion of the work force may be needed in supervisory positions. To reduce turnover, which disrupts the production process, high wage firms will substitute future benefits in the form of pensions for current benefits in the form of wages. Discontinuities in production are also reduced through vertical integration.

Information about low and high paying firms is important for another reason besides the light it sheds on how production is organized in these two types of firms. Since low paying firms are a source of employment for workers with relatively low productivity, it is of some interest to inquire into the extent

to which low pay among workers is attributable to their employment in such firms. In approaching the question of why some workers are paid less than others in this manner, low wage employers can be viewed as providing employment opportunities with attendant low earnings, not because they discriminate against certain groups of individuals, but because the production processes that are most efficient for their mode of operation do not require high quality labor and, furthermore, they inhibit their paying high wages.³³

A procedure for verifying this view would be to sector firms according to whether they are low paying or high paying.³⁴ With this sectoring of firms, one would expect, as indicated above, that the mix of workers and capital is dissimilar between the two sectors. Assuming this is so, to what extent are differences in individual earnings in low and high paying firms due to the characteristics of the workers and capital employed in each type of firm? Also, to what extent are workers with similar characteristics remunerated in the same way in each type of firm?

One way of answering the first of these questions is to separately estimate wage rate equations for workers in low and high paying firms. The variables in each equation would reference the quality of labor and the quantity and quality of capital, as well as other variables controlling for occupation, industry, geographical location, union status, etc. Earnings differences between each type of firm due to factors other than labor quality can then be estimated by assuming that the workers in low paying firms have the same characteristics as workers in high paying firms. Additionally, earnings differences between each type of firm due to factors other than the quantity and quality of capital can be estimated assuming that the quantity and quality of capital in low paying firms is the same as in high paying firms. Subtracting the sum of these

differences from the total differential in earnings of workers in high and low paying firms yields an estimate of the earnings discrepancy which is due to the differential rate of return to labor and capital in both sectors.³⁵

The answer to the second question posed above is obtained by determining which coefficients of the variables referencing labor and capital are significantly different from zero (and have the right sign) in each sector, and where both coefficients for a given variable are significant, whether they are significantly different from each other. For example, it may turn out that being female or black has no effect on earnings in low paying firms but both groups earn less than their white male counterparts in high paying firms. Were this outcome observed, one could then go on to estimate the amount by which the earnings of these groups would rise in high paying firms if they had been paid at the same rate as white males in high paying firms.

The primary point to be emphasized by this discussion is that information about firms, in particular, whether they are low or high paying, adds an important additional dimension in assessing how earnings outcomes are determined in the labor market.

D. Measuring the Effects of Minimum Wage Legislation

Economists have long been interested in government policies which are believed to impose restrictions on the free operation of labor markets. Minimum wage legislation falls into this category. As a result of amendments to the Fair Labor Standards Act (FLSA) in 1977, the federal minimum wage rose from \$2.65 per hour in 1978 to \$3.35 per hour in 1981; the same minimum wage prevails today. As of September 1982, 85.5 percent of all private sector nonsupervisory employees were covered under the 1977 amendments; the corresponding figure for all employed wage and salary workers was 63.5 percent.³⁶

be used in policy analysis, since, once again, consideration is being given to lowering the minimum wage for young people. The argument for a youth differential is based on the supposition that the minimum wage has led employers to substitute adult workers for younger ones, and a lower minimum wage for youths would result in more jobs being made available to them. To test this proposition, Cotterill and Wadycki (1976) estimated the percentage of workers employed in establishments with annual sales of \$300,000 or more in 8 retail trade industries in 31 Standard Metropolitan Statistical Areas (SMSAs).³⁹ This percentage was used to proxy the extent to which workers in each industry and SMSA were covered by the Federal minimum wage. The need for this proxy stems from the absence of information in their data set which would permit identification of individuals employed in covered establishments.

Cotterill and Wadycki tested two hypotheses: 1) the wage rate of retail trade employees is higher in SMSAs where minimum wage coverage is greater and 2) firms in SMSAs where minimum wage coverage is greater compensate for higher wage rates by employing more adults and fewer youngsters, i.e., higher quality labor.⁴⁰ The first hypothesis is tested using a wage model which excludes personal characteristics variables; not surprisingly, the hypothesis is consistent with the data. The second hypothesis is tested by adding personal characteristics variables to the original wage model. If employers in high coverage SMSAs substitute adult workers for younger ones, the higher wage in these areas would be "picked up" by the personal characteristics variables, thereby reducing the coefficient of the coverage effect variable in the modified model. No reduction in the coverage effect variable was observed, suggesting that young people had not been replaced by older workers when the minimum wage was extended to retail trade.

By raising the wage above that which would prevail in a competitive labor market, workers whose marginal revenue product is less than the minimum wage are subject to disemployment. With respect to the firms in which such workers are found, the disemployment effect may be so severe that they will cease operation. In general, all else being the same, the disemployment effect will be greater the lower a firm's average wage rate. Moreover, the adverse impact of the minimum wage may be greater for both low wage and high wage workers in low wage firms than in high wage firms, since the minimum wage impacts on the firm, and only indirectly affects individual workers through changes in firm behavior. As a result of the minimum wage, some high wage workers in low wage firms may become disemployed if total firm employment shrinks. On the other hand, if the proportion of low wage workers is small in a high wage firm, there may be no or only a small disemployment effect.

The earliest studies of the effects of minimum wage legislation focused on low wage industries, since such effects are most reliably detected when a significant proportion of an industry's work force is comprised of low wage workers. These studies attempted to isolate the impact of the minimum wage by comparing changes in employment before and after imposition of (or an increase in) the minimum wage among a test group and control group of firms, e.g., between covered and noncovered firms.³⁷ While this approach is the most direct one in getting at the impact of minimum wage legislation, the data that have been collected suffer from several deficiencies. Among these deficiencies are 1) retrospective data have been utilized so that firms that closed down between the time legislation was initiated and a survey taken were omitted from the studies, thereby understating the adverse impacts of the minimum wage, 2) the studies have typically

measured employment only in terms of numbers of workers with no correction for hours worked, again, possibly understating adverse impacts, 3) inclusion of workers, such as professional workers and managers, who are normally exempt from the FLSA makes interpretation of the data difficult, and 4) it has not been possible to control for prior employment trends.

The need to control for prior employment trends is due to the implicit assumption that in the absence of minimum wage legislation, the test and control groups would grow at the same rate.³⁸ For example, assume that it is found that employment in retail industries characterized by low wage firms grew more rapidly than retail industries characterized by high wage firms, despite extension of the FLSA to the former group. In this case, if employment in the former group had been growing even faster than in the latter group prior to the extension of coverage, it would be incorrect to infer that no disemployment had occurred.

Because of the aforementioned difficulties, industry studies of the disemployment effect of the minimum wage have fallen out of vogue. Were a SIPP-ED file in place, each of the problems just noted could be readily resolved.

In recent years, investigators concerned with minimum wage issues have turned their attention to groups in the population with specific demographic characteristics. The group receiving the most study is that of young people; most of the time-series studies have been confined to this group. More recently, cross-sectional data have been utilized to study the effects of minimum wage legislation, holding constant factors which cannot be controlled for in the time-series data.

The study of the effect of the minimum wage on the substitutability of adult labor for that of youths is particularly insightful as to how a SIPP-ED file can

It should be clear from this brief review of Cotterill and Wadycki's study that a much simpler and more direct test of their hypotheses would have been possible if a SIPP-ED file had been available to them. The SSEL portion of this file would have permitted them to identify firms and workers covered by the federal minimum wage law, thereby vitiating the need to develop a proxy variable based on geographical variations in coverage. Moreover, it would have eliminated the need for an indirect test of the substitution effect of the minimum wage. By grouping workers in covered and in uncovered firms in various industries, and controlling for other factors that govern the ratio of young to adult workers, e.g., the geographical and occupational distribution of an industry's work force, variations in the youth/adult ratio attributable to employment in firms covered by the federal minimum wage law could be directly ascertained.

It should be noted that in studying the impact of the extension of Federal minimum wage coverage on employment in say, retail trade, minimum wage affects should be distinguished from size of firm affects. For example, assume that after the extension of the Federal minimum wage in retail trade it was found that the youth/adult worker ratio was lower in SMSAs where the coverage was high than in SMSAs where it was low (or lower in covered than noncovered firms were such data available). One could not infer from these data alone that as a result of the extension of coverage adult workers were substituted for younger ones, since covered firms are larger in size than noncovered firms and would tend to have a lower youth/adult worker ratio even in the absence of the minimum wage. This problem is amenable to analysis using a SIPP-ED file since it would contain information on size of firm; hence, one could control for variations in this variable.

Cross-sectional data have also been used by Leighton and Mincer (1981) in assessing the impact of the minimum wage on OJT. Since OJT is part of the total compensation package, any exogenous increase in the wage rate, in this case, the minimum wage, should restrict the amount of OJT that low wage firms can provide.

Given that the Federal minimum wage is uniform across states, the authors estimate the level of wages in each state for workers with the same characteristics. By identifying low wage and high wage states in this manner, it can be assumed that, on average, firms in low wage states pay lower wages than firms in high wage states. Holding the proportion of workers covered by Federal minimum wage legislation in a state constant, the lower the value of the state wage, the greater the negative impact of the minimum wage,⁴¹ i.e., the less likely are workers to have participated in OJT. Leighton and Mincer measure OJT directly from responses indicating whether a person received training in his or her current job and indirectly from the change in an individual's wage and from their length of job tenure. The empirical data suggest that, indeed, the minimum wage tends to discourage OJT.⁴²

As before, availability of a SIPP-ED file would provide a means of getting directly at the effect of the minimum wage on OJT, since the average wage paid by a firm is a datum in this file; hence, there would be no need to estimate state wage proxies. Additionally, one can control for the characteristics of the firms themselves to explore how OJT is related to the production process.

Besides the possibility that the minimum wage may affect the amount of OJT that a firm offers, it may also affect a firm's ability to provide health insurance and retirement fringe benefits, since a wage floor limits the trade-off between wages and other forms of compensation. Here again, a SIPP-ED file could be utilized to assess still another aspect of the economic consequences of the minimum wage.

E. Structural Unemployment

An important issue in maintaining full employment is the proper mix of policies to meet the challenge of structural change. The direct effects of structural change, whether arising from the introduction of new technologies, the substitution of foreign for domestic output, or long-term shifts in consumption patterns, are typically localized to firms producing a particular line of products or to specific areas in which firms are engaged in a variety of activities. In either instance, the structural disequilibria are such as to result in relatively large changes in firm employment, often of such magnitude that firms are forced to close their doors. One recent estimate places the annual number of jobs lost due to major plant closings between 1978-1982 at 900,000.⁴³ The essential feature of these disequilibria is that the period of time needed before they work themselves out is longer than the standard business downturn.

An issue of long standing is what happens to workers who are displaced from their job as a result of structural disequilibria. How long do they remain unemployed vis-a-vis other workers who separate from an employer? What sources of income, including cash and noncash government transfers, do they draw on when they are unable to find work? When they find a job, how do earnings in the new job compare to earnings in the old one? If there is an earnings loss, how much of this loss is recouped, say, after 2 years?

A major problem in answering these questions is that workers do not know if they are structurally unemployed. One way of identifying such workers is to ascertain what has happened to the firms in which they were last employed.

If the firm has undergone a substantial decline in employment or has closed down for a relatively long period of time, say, longer than the typical recession, one may presume that it has undergone a shock which is typical of the shocks experienced by firms subject to structural disequilibria. It also can be presumed that the employees of these firms experience the aftereffects of such shocks. For some, the aftereffect is loss of a job. For others who are able to retain their job, the aftereffect may be reduced earnings or diminished pay raises instead of actual wage cuts. However, just how poorly workers affected by structural disequilibria fare, relative to job changers and job stayers in firms where demand conditions are stable, is not known.

As indicated previously, the SSEL and LED contain longitudinal data on employment and operational status. A SIPP-ED file would enable one to determine the extent to which firms are subject to severe, long-term shocks as evidenced by plant closures and substantial reductions in employment, and how such shocks affect their work force.

F. Identifying High Tech Workers and High Tech Firms

Despite the importance of new technologies for improving productivity, regaining our competitive advantage in international markets, and maintaining our defense posture, there is no widely accepted definition of a high tech industry. Using 3 different definitions, Riche, Hecker, and Burgan (1983) estimate that from 2.8 to 13.4 percent of all wage and salary workers were employed in high tech industries in 1982.⁴⁴ The first figure is based on a definition which includes industries with an R and D to net sales ratio of at least twice the average for all industries. The second is based on a definition which includes industries with a ratio of technology-oriented workers⁴⁵ to all workers of at least 1.5 times the industry-wide average.^{46, 47}

High tech industries have been cited as having a large group of high and low wage workers whereas other industries are comprised of workers who are concentrated in the middle of the earnings distribution. While it is useful to know how workers in high tech and other industries differ and the differential growth of employment in the two kinds of industries, it is equally important to know the characteristics which differentiate high tech from other firms and the differential in the rate of growth of the two types of firms.

As is self-evident, not all firms in high tech industries utilize the latest technology, and new techniques of production are utilized by firms in industries besides those labeled as high tech. One approach to distinguishing between the two types of firms would be to compare the characteristics of the industries denoted on a priori grounds as high tech with other industries and then to use this information to identify high tech firms.⁴⁸ To illustrate this approach, assume that the a priori criterion used to denote high tech industries is one of the definitions noted above, namely, that the ratio of high tech to all workers in a given industry to the similar ratio for all industries is higher than some minimum value. Assume also that the high tech industries exhibit high values of the following ratios: capital expenditures for new computers to all capital expenditures, capital expenditures to asset value, and capital to labor. Given a set of characteristics which permit the bifurcation of industries, the multivariate technique of cluster analysis can then be applied to identify high tech firms within both high tech and other industries.

Cluster analysis is a way of analyzing multivariate data.⁴⁹ It is particularly useful in creating a classification system in that it enables one to group observations, in this case, individual firms, into homogeneous classes or clusters

without imposing a priori specifications on the data other than the choice of variables to be used in the analysis. Using cluster analysis, each observation can be compared with every other observation and a measure of distance can be computed for every matched pair. Grouping the two observations with the smallest distance between them, the number of observations is reduced by one; the process is then repeated. The end result is two or more clusters, determined by the data, where each cluster of firms represents a homogeneous set of observations.

The outcome of the cluster analysis is a partitioning of firms into categories, say, high tech and nonhigh tech firms. An advantage of applying the aforementioned two-stage procedure using a SIPP-ED file is that it provides an independent test of how well the procedure works. For if the approach is successful, the proportion of workers who are technology-oriented among the firms classified as high tech (taken as a group) will be higher than the similar proportion for firms classified as nonhigh tech (again, taken as a group), and the difference in proportions will be greater than the corresponding difference when industries are classified as high tech and nonhigh tech. An additional advantage of the SIPP-ED file is that industries can be disaggregated to the 4-digit SIC level. At this level of detail, a better determination can be made of the variables to be included in the cluster analysis than at the more aggregated census industry level.

Having identified high tech firms, in contrast to high tech industries, insights can then be obtained as to how production processes in these firms differ from their nonhigh tech counterparts. At the same time, it will enable one to better define high tech occupations and how workers in these (and other) occupations in high tech firms differ from similar workers in nonhigh tech firms.

G. Implications of the Transition from a Goods to a Service Economy

One of the most striking changes in the U.S. economy since World War II has been the growth of service sector employment. Between 1950-1983, the annual rate of growth in goods producing industries has averaged 1.0 percent. In service industries, however, the rate of growth has averaged 3.1 percent per year over the same period. This rapid growth, coupled with the fact that 1 out of 2 wage and salary workers in the service sector is employed in enterprises with less than 100 workers, compared with 1 out of 4 in the goods sector, suggests that small firms are becoming a more common feature of the economic landscape.⁵⁰

Although investigators have used size of firm as a variable in labor market studies, information on how small and large businesses differ with respect to the kinds of workers they employ, the economic rewards which individuals receive when working in various size firms, and other factors that distinguish the two size classes is sparse. Information about how small and large firms differ in the goods and service sector is even scantier.

It is known that small firms employ a larger fraction of younger and older, female, and part-time workers among their work force than large firms (Barth, Cordes, and Haber, 1984). And as indicated above, small firms pay less than large ones, holding the characteristics of workers constant. The incidence of fringe benefits, particularly private pension benefits, is also lower among small firms (Mellow, 1982a). It is not clear, however, whether these differences apply equally to all industries or are substantially larger among service industries than goods industries.

Among the ways in which goods and service firms differ, besides the characteristics of their work force, is in their capital/labor ratio. For firms of the same size, the capital/labor ratio is lower in the service sector than in the goods sector. Additionally, the productivity of a worker in the service sector is likely to be less than that of a similar worker in the goods sector working with the same amount of capital. To the extent that this is so, profit margins may be lower in the service sector, and this may explain why service sector employees earn less. From another perspective, given that the requirement for physical capital in a typical service sector firm is lower than that of its counterpart in the goods sector, the changing distribution of firms between the two sectors may also mean a reduced demand for reproducible capital.⁵¹ On the other hand, the shift to service sector employment has led to an increase in demand for workers who have invested in human capital.

It has been suggested by some that shifts in employment to high tech industries within the manufacturing sector and from the goods to the service sector is resulting in a decline in the middle class (Thurow, 1984). The high tech industries, it is said, are composed of high paid professional workers and low paid assembly workers in contrast to the "smokestack" industries in which highly paid skilled workers and almost as highly paid operatives predominate. Likewise, high and low paying jobs are believed to be more characteristic of the service sector than the goods sector. This may be due, in part, to the greater prevalence of small firms in the former sector.

It has also been observed that the distribution of earned income, i.e., wages and salaries plus income from self-employment, has become more unequal over

time for men but has remained about constant for women (Henley and Ryscavage, 1980). The sectoral shifts just noted are consistent with these trends. Given the posited distribution of earnings in high tech and service industries, a relative shift of employment among males to these industries would result in a more unequal distribution of male earnings. On the other hand, women have been employed in service industries for decades, so that the expansion of jobs in this sector (and similar low paying jobs in high tech industries) would have little effect on the distribution of their earnings.

As indicated, small (large) firms are characteristic of the service (goods) industries. To the extent that both small and large firms pay less in the service sector than in the goods sector, when men shift from the latter to the former sector there is a tendency for them to "slide down" the income distribution, even when they find employment in the same size firm. This same tendency may be absent for women, since a disproportionate fraction of women are employed in the service sector.

A SIPP-ED file would provide a basis for measuring the effect on the earnings distribution of shifts in employment within and between the two sectors and different types of firms, for both men and women. The advantage of using this data set is that it enables one to get at the process by which the earned income distribution changes over time. One way in which this process can be quantified is to sector the economy into high tech and nonhigh tech firms within both the goods and service sectors, and then to further stratify each group into small and large firms.⁵² For each of these groups, one can compute a measure of inequality of earned income and obtain a weighted inequality measure over all groups taken as a whole.⁵³ By comparing the measure of inequality and its

components for successive SIPP panels, the effects of diverse structural changes impacting simultaneously on the economy can be separately determined. With a decomposable measure of earnings inequality, one can also partition each sub-sector by personal characteristics, e.g., age, sex, race, and marital and head of household status, and compare the affects of changes in demographic and economic variables. A similar analysis can be performed for individual SIPP panels; in this case, changes in the distribution of earned income would occur within an approximately closed population, since the composition of any given panel, once determined, is essentially fixed.⁵⁴

3. The Pilot Study

A principal part of the pilot study is designed to assess the availability, sources, coverage, and content of the various economic data files maintained by the Bureau of the Census and to explore study areas and issues to which a data set combining micro-worker and firm data would be applied. In the course of this study, specific demographic and economic variables have been identified which should be incorporated into such a data set. Additionally, it was anticipated that methodological problems inherent in this undertaking would be revealed; indeed, this has been the case.

A second phase of the pilot study is to investigate the efficiency of four alternative methods of identifying an individual's employer. Each method is based on different information for searching the SSEL and identifying the employer's census file number (CFN). The first utilizes information on employer name, the state of residence and/or zip code of the employee, and census industry code. The same information is used in the second method; however, additional reference

materials, e.g., 1980 Census Company Name and Place of Work lists, Dun and Bradstreet reference books, Standard and Poor directories, and telephone books, will be used to obtain the exact address of an individual's employer. The third method uses the employer's name and exact address if known. In the last method, if the EIN is known, it is used in conjunction with the information available in the first three methods to identify the employer's CFN. For each method, match rates and cost information will be developed for a small sample of workers.

A third phase of the study is the construction of a pilot SIPP-ED file in which the SIPP portion of the file would be restricted to full-time workers in large manufacturing establishments; the source of the economic data would be the LED. The objective in this phase is to calculate match rates between workers in SIPP and their establishments in the LED.

Given the importance of the wage determination process, one of the areas noted above, e.g., the relationship between wage rates and capital or low wage workers and low wage firms, would be studied when the pilot work file is completed. Demonstration of the utility of this research endeavor in terms of its contribution to the economic literature will constitute the final phase of the pilot study.

4. Methodological Problems in Matching Demographic and Economic Data

In this section, attention is focused on two methodological problems. One is central to the development of a SIPP-ED file; the other is peripheral but places a constraint on the way in which the file can be applied. The first problem deals with procedures for tying workers to their establishment and company. The second relates to the estimation of data, which although available for large establishments and companies, are generally not collected for small ones.

B. Estimating Missing Economic Data for Small Establishments and Companies

Some economic data, in particular, the gross value of depreciable assets and supplemental labor costs, are collected for some small establishments and companies only when they happen to be included in annual surveys and/or special reports. Given that the primary objective of the Bureau of the Census in collecting economic data is the measurement of industry output, and that the bulk of output in any industry is produced by large establishments and companies, the quantity and quality of the data that is collected for these businesses need not be as good as that for their larger counterparts. This is particularly so with respect to accounting data, such as asset information, where accounting practices may vary from firm to firm.⁵⁷ Even if substantial improvements were made in estimating missing economic data for small establishments and companies, these would translate into only small improvements in estimating industry aggregates. Hence, while there is interest in improving the economic data for firms of all sizes, to do so by collecting additional data from small firms would require a disproportionately large outlay of resources.

As indicated, information on assets is not generally available for small establishments. In the CM (and LED), the asset values for such establishments are imputed by multiplying their value of shipments by the average capital/value of shipments ratio for large establishments in the same 4-digit SIC class (as calculated from the ASM).⁵⁸ Despite the fact that asset information is not collected for many of the firms in which individuals work, the use of an economic model, including industry, firm size, and other variables, may enable one to obtain reasonably accurate estimates of capital for small establishments.

Economic theory suggests a number of relationships which influence the amount of capital that a firm employs in its production process. In particular, since capital intensity varies with establishment size in closely related industries, it seems reasonable to assume that information about the number of employees in an establishment can be used to further refine estimates of its capital assets. All else being the same, one would expect the smaller an establishment, the lower would be its capital/labor ratio. Additionally, holding everything else constant, including establishment size, low wage establishments will substitute labor for capital in order to economize on the use of the relatively expensive factor, i.e., capital. Thus, low wage establishments will tend to have a lower capital/labor ratio than high establishments.

Even among establishments of the same size whose wage rate is also the same, one would expect a lower capital/labor ratio, the higher the proportion of production workers among all workers. As mentioned, when the proportion of production workers among all workers is high, or conversely, when the percentage of workers who supervise production is low, this comes about because a firm has few assets, relative to labor, to monitor. Additional relationships between assets and other variables may exist. For example, it may be that newer establishments in an industry are more capital intensive than older ones; likewise, regional variations in entrepreneurial ability may give rise to corresponding variations in capital intensity.

Besides economic relationships, engineering relationships also may be useful in estimating capital intensity. For example, it is plausible that an establishment's capital/labor ratio is positively related to purchased electricity per employee; while the former can be derived only for small establishments in the ASM, the latter is available for all establishments in the CM.

It will be noted that the dependent variable in our economic model is the capital/labor ratio rather than the capital/value of shipments ratio. When firms change their level of output, they do so by adjusting overtime hours as well as the number of workers in their employ. To the extent that overtime hours is chosen to adjust labor inputs, variations in employment over the business cycle will be smaller than the corresponding variations in value of shipments. For this reason, where the focus of inquiry is specifically on micro-estimates of firm capital, use of the capital/labor ratio as a multiplicative factor will yield more stable estimates and, hence, is preferred to the capital/value of shipments ratio.⁵⁹

An assessment of the utility of an economic model to estimate the assets of an establishment can be made using the ASM. As mentioned, approximately four-fifths of the ASM is comprised of small establishments. Given the large size of this sample, the economic model can be fitted to establishments in closely related industries, omitting a subset of observations which could be used as a "live" test of how well the model performs against the current procedure. Such a test should be restricted to establishments with information as originally reported by respondents to avoid cases where reported values have been computer or analyst corrected to conform to Bureau of the Census editing procedures.

Finally, an economic model can also be utilized to estimate fringe benefits for small establishments and small companies. It is plausible to assume that fringe benefits in a firm are related to its size, average wage level, legal form of organization, industry, and region where it is located. With a SIPP-ED file, more refined estimates of fringe benefits per employee can then be obtained by taking account of the percentage of employees who are covered by life and medical

insurance and a private pension plan in a given group of firms, say, (small) high paying establishments in manufacturing. Given this information, the average value of these benefits per covered and noncovered worker can be calculated for each establishment in the group. The same procedure can also be used to estimate fringe benefits per covered and noncovered worker in large establishments and companies, omitting the first step in the procedure, i.e., the development of an economic model to estimate fringe benefits, since these benefits can be obtained directly from the economic data for large firms. With appropriate information in SIPP, these estimates could provide a basis for imputing an important component of private noncash benefits to individual workers. Although it should be evident from the discussion of this paper, this last illustration is indicative of the benefits to be derived from a SIPP-ED file.

FOOTNOTES

- 1/ These industries are denoted by the Bureau of the Census as in-scope industries. Out-of scope industries, with respect to the economic censuses, include transportation, communications, and public utilities; finance, insurance, and real estate; and some service industries. The ES data for 1982, however, will also cover agriculture.
- 2/ Nelson, McMillen, and Kasprzyk (1983), pg. 1.
- 3/ Ibid., pp. 4-6.
- 4/ Ibid., pp. 7-9.
- 5/ Ibid., pp. 9-10.
- 6/ For a discussion of the measurement of labor force and employment status based on SIPP data, see Ryscavage (1984).
- 7/ For ease of exposition, the terms company and enterprise are used interchangeably. Both terms reference a parent company, comprised of one or more establishments and/or subsidiaries, which is a completely independent business organization.
- 8/ The alpha number also identifies a parent company in the ES and, hence, is the link for integrating the economic data in the SSEL for establishments with similar data in the ES for the parent company.
- 9/ In contrast to its mailing address which may be a post office box number.
- 10/ The CFN is composed of 10 digits. For single establishment enterprises, the CFN is the company's EIN preceded by a zero. For establishments of enterprises with two or more establishments, the CFN is the parent company's 6 digit alpha number followed by a four digit establishment number. Establishments in the LED are identified by their CFN.
- 11/ Economic census years only and only for businesses identified by an EIN in the EC-EI file
- 12/ A more accurate estimate of the average wage can be obtained by correcting the earnings data to take account of the hours usually worked in the 1st quarter in the industry in which the employer is engaged. This latter figure can be derived from SIPP.
- 13/ Additional information about the SSEL is found in Bureau of the Census (1979).
- 14/ Approximately 22 percent of the ASM sample falls into this size group.
- 15/ In 1982, CM questionnaires were sent to approximately 350,000 establishments.

- 16/ Value added is measured by the difference between the total value of shipments and the cost of materials and services plus the net change between beginning and end-of-year inventories.
- 17/ Additional information about the LED is found in Monahan (1983).
- 18/ Additionally, the ES data include in-scope establishments of companies primarily engaged in out-of-scope industries, e.g., retail appliance stores of a public utility are included in the ES.
- 19/ Single establishment companies and multi-establishment companies all of whose establishments are engaged in a single 4-digit SIC industry are classified as single industry enterprises. All other companies are denoted as multi-industry companies. From payroll data, the 4-digit SIC of companies in the latter category are obtained by first ascertaining their largest census industry division (e.g., manufacturing, retail trade) and then determining their primary 4-digit SIC industry within that division.
- 20/ For mineral and manufacturing industries, employment is given by the average of paid production workers for the pay periods including the 12th of March, May, August, and November plus all other paid employees in the pay period including March 12; for construction, employment is given by the average of all paid workers in the pay periods including the 12th of March, May, August, and November; for all other industries, employment represents the number of employees on the payroll in the pay period including March 12.
- 21/ For companies with fewer than 500 employees classified in the mineral, construction, and manufacturing industries, capital expenditures are only cumulated for their operating establishments in these three industries.
- 22/ One reason for more efficient, larger firms to acquire higher quality labor is that potential losses in output, when production is disrupted due to worker absence, negligence, or error, are greater than for less efficient, smaller firms. Hence, it pays the former to incur monitoring costs or to attempt to defray them by hiring higher quality labor (Oi, 1983).
- 23/ Indeed, the difficulty of screening applicants has given rise to the theories of signalling (Spence, 1974) and of statistical discrimination (Phelps, 1972).
- 24/ Of interest, when Brogan and Erickson included industry concentration and establishment size variables in their wage equation along with a capital per establishment variable (based on industry data), they found that while the last variable remained highly significant neither the concentration or size variables were statistically significant.
- 25/ Kwoka (1983) summarizes some of the major wage rate studies in which industry information on plant size is utilized. In the seven studies cited in which plant size is a variable entering the analysis, it has a positive and significant effect on wages in every one.
- 26/ Some evidence for this latter proposition, based on a 1968 BLS survey of 1,149 establishments, is found in Bailey and Schwenk (1971).

- 27/ Kwoka (1983).
- 28/ On the other hand, young firms may have more modern equipment which would contribute to higher productivity and earnings.
- 29/ In 1977, the last year for which data are available, fringe benefits accounted for 15.5 percent of total compensation (Bureau of Labor Statistics, 1980).
- 30/ Doeringer and Piore (1971).
- 31/ Department of Commerce (1984).
- 32/ Minimum Wage Study Commission (1981).
- 33/ The rapid growth in labor force participation among women and their absorption by low paying firms is consistent with the puzzling phenomenon of why the female/white male wage rate failed to decline during the decade of the 1970's. See Green (1984).
- 34/ As an example, a low paying firm might be defined as one with an average wage rate which is less than one-half the median of all firms; this criterion is similar to one proposed by Fuchs (1967) in defining the poverty level of income. A desirable feature of this criterion is that it focuses attention on the distribution of firms in terms of the wage they pay their employees. Obviously, other criteria can be used as well, in particular, firms can be classified by wage interval. We assume that firms are categorized in this manner; for simplicity, two categories are used and are denoted as "low" and "high" paying firms.
- 35/ Similar estimates need to be made assuming that the quality of workers and the quantity and quality of capital in high paying firms are the same as in low paying firms.
- 36/ Department of Labor (1983).
- 37/ See, for example, Department of Labor (1959).
- 38/ Brown, Gilroy, and Kohen (1982).
- 39/ At the time of their study, retail establishments were covered by the Federal minimum wage if they had annual sales of \$250,000 or more and were part of an enterprise with annual sales of \$ 500,000 or more (Cotterill and Wadycki, 1981, p. 84).
- 40/ Cotterill and Wadycki consider state as well as Federal minimum wage affects. For simplicity, and with no loss in generality, the discussion in the text is restricted to the latter.

- 41/ See also Welch (1974) where this approach to measuring the impact of the minimum wage is developed in greater detail.
- 42/ Using a different method to estimate a minimum wage measure, Hashimoto (1982) reaches the same conclusion.
- 43/ Bluestone, Harrison, and Gorham (1984).
- 44/ For other approaches to defining high tech industries, see Department of Commerce (1983).
- 45/ Defined as engineers, life and physical scientists, mathematical scientists, engineering and science technicians, and computer specialists.
- 46/ Of the approximately 12.3 million wage and salary workers in high tech industries under the first definition, 7.5 million were employed in manufacturing. Of the remaining 4.8 million workers, 2.0 million were employed in two industries -- wholesale trade, machinery equipment and supplies and heavy construction, except highway and street.
- 47/ Under a third definition, which is a composite of the two noted in the text, 6.2 percent of all wage and salary workers were employed in high tech industries in 1982.
- 48/ To simplify the problem, the universe of industries might be restricted at first to those in manufacturing.
- 49/ For a discussion of clustering techniques, see Hartigan (1975).
- 50/ For a discussion of the growth of small firms and the relationship between such growth and that of total employment, see Birch (1979).
- 51/ The demand for capital depends on the number of firms as well as the capital requirement of the typical firm in each sector. Since service sector employment is growing relative to employment in the goods sector, our conclusion concerning the demand for capital is, at best, a tentative hypothesis.
- 52/ In the absence of being able to identify high tech and nonhigh tech firms, high tech and nonhigh tech industries can be distinguished. To simplify the analysis even further, only four sectors might be used, i.e., small and large firms within the goods and service sectors.
- 53/ To do this, it is necessary to use a measure of income inequality which is decomposable; hence, a measure other than the Gini coefficient is required, e.g., Thiel's entropy measure (Thiel, 1967), pp. 94-96.

- 54/ Even within a SIPP panel, some individuals will enter or reenter the labor force while others leave or retire from the labor force. To obtain a completely closed population, only individuals who were employed in both the beginning and ending reference periods should be considered.
- 55/ In cases where more than one activity is located at a given physical address e.g., a sales activity and manufacturing activity may be located on different floors of the same building, the census industry code can be used to determine in which activity an individual works.
- 56/ In the Income Survey Development Program somewhat more than 50 percent of the respondents used their W-2 forms to provide income information.
- 57/ Because of the difficulty of collecting asset information, it was only in 1977 that information about beginning-of-year assets and retired assets was asked in the ASM. Prior to 1977, only information on end-of-year assets was collected. With the additional information now collected, it is possible to determine the process by which assets increase or decrease from one year to the next. It is also easier to detect errors in end-of-year estimates of assets.
- 58/ The approach is used because in some 4-digit SIC classes the number of small establishments and their response rate is low. It is felt that the average capital/value of shipment ratio for large establishments in a 4-digit SIC class is more reliable than the corresponding value for small establishments.
- 59/ A similar argument is made by Klotz, Madoo, and Hansen (1980) for defining establishment size in terms of assets rather than value of shipments. On the other hand, because some small high tech establishments can account for a large share of an industry's output, it may be desirable in some cases to define establishment size in terms of value of shipments.

REFERENCES

- Antos, Joseph, "Union Effects on White-Collar Compensation," Industrial and Labor Relations Review, April 1983, pp. 461-479.
- Bailey, William and Schwenk, Albert, "Wage Differences Among Manufacturing Establishments," Monthly Labor Review, May 1971, pp. 16-19.
- Barth, James; Cordes, Joseph; and Haber, Sheldon, Worker Characteristics and Size of Business: Are there Systematic Empirical Relationships? U.S. Small Business Administration, Office of Advocacy, 1984.
- Birch, David, The Job Generation Process, M.I.T. Program in Neighborhood and Regional Change, Cambridge, Mass., 1979.
- Bluestone, Barry; Harrison, Bennett; and Gorham, Lucy, Storm Clouds on the Horizon: Labor Market Crisis and Industrial Policy, Economic Education Project, Brookline, Mass., May 1984.
- Borjas, George, "Job Mobility and Earnings over the Life Cycle," Industrial and Labor Relations Review, April 1981, pp. 365-376.
- Brogan, Robert and Erickson, Edward, "Capital-Skill Complementarity and Labor Earnings," Southern Economic Journal, July 1975, pp. 83-88.
- Brown, Charles; Gilroy, Curtis; and Koben, Andrew, "The Effect of Minimum Wage Legislation on Employment and Unemployment," Journal of Economic Literature, June 1982, pp. 487-528.
- Bureau of the Census, The Standard Statistical Establishment List Program, Technical Paper 44, January 1979.
- Bureau of Labor Statistics, Employee Compensation in the Private Nonfarm Economy, 1977, Summary 80-5, April 1980.
- Cotterill, Philip and Wadycki, Walter, "Teenagers and the Minimum Wage in Retail Trade," Journal of Human Resources, Winter 1976, pp. 69-85.
- Department of Commerce, International Trade Administration, An Assessment of U.S. Competitiveness in High Technology Industries, February 1983.
- Department of Commerce, Characteristics of the Population Below the Poverty Level: 1982, Current Population Reports, Series P-60, No. 144, February 1984.
- Department of Labor, Wage and Hour and Public Contracts Division, Studies of the \$1 Minimum Wage, 1959.
- Department of Labor, Employment Standards Administration, Minimum Wage and Maximum Hours Standards Under the Fair Labor Standards Act - 1983 Report, 1983.

REFERENCES (cont'd)

Doeringer, Peter and Piore, Michael, Internal Labor Markets and Manpower Analysis, D.C. Heath and Co. 1971.

Fuchs, Victor, "Redefining Poverty and Redistributing Income," The Public Interest, Summer 1967, pp. 89-95.

Green, Gordon, Wage Differentials for Job Entrants, by Race and Sex, The George Washington University, Ph.D. Dissertation, 1984.

Hartigan, John, Clustering Algorithms, John Wiley and Sons, 1975.

Hashimoto, Masanori, "Minimum Wage Effects on Training on the Job," American Economic Review, December 1982, pp. 1070-1087.

Henley, Peter and Ryscavage, Paul, "The Distribution of Earnings Among Men and Women," Monthly Labor Review, April 1980, pp. 3-10.

Klotz, Benjamin; Madoo, Ray; and Hansen, Reed "A Study of High and Low 'Labor Productivity' in U.S. Manufacturing," in John Kendrick and Beatrice Vaccara, eds., New Developments in Productivity Measurement Analysis, National Bureau of Economic Research, University of Chicago Press, 1980, pp. 239-292.

Kwoka, John, "Monopoly, Plant, and Union Effects on Worker Wages," Industrial and Labor Relations Review, January 1983, pp. 251-257.

Leighton, Linda and Mincer, Jacob, "The Effects of Minimum Wages on Human Capital Formation," in Simon Rottenberg, ed., The Economics of Legal Minimum Wages, American Enterprise Institute for Public Policy Research, Washington, D.C., 1981, pp. 155-173.

Masters, Stanley, "Wages and Plant Size: An Interindustry Analysis," The Review of Economics and Statistics, August 1969, pp. 341-345.

Mellow, Wesley, "Determinants of Health Insurance and Pension Coverage," Monthly Labor Review, May 1982, pp. 30-31.

Mellow, Wesley, "Employer Size and Wages," The Review of Economics and Statistics, August 1982, pp. 495-501.

Mellow, Wesley and Sider, Hal, "Accuracy of Response in Labor Market Surveys: Evidence and Implications," Journal of Labor Economics, October 1983, pp. 331-344.

Mincer, Jacob and Jovanovic, Boyan, "Labor Mobility and Wages," in Sherwin Rosen, ed., Studies in Labor Markets, National Bureau of Economic Research, University of Chicago Press, 1981.

Minimum Wage Study Commission, "A Demographic Profile of Minimum Wage Workers," Report of the Minimum Wage Study Commission, Chapter 1, May 1981.

REFERENCES (Cont'd)

Monahan, James, Procedures for Using the Longitudinal Establishment Data File, Bureau of the Census, April 1983.

Nelson, Dawn; McMillan; David; and Kasprzyk, Daniel, An Overview of the Survey of Income and Program Participation, Bureau of the Census, December 1983.

Oi, Walter, "Heterogeneous Firms and the Organization of Production," Economic Inquiry, April 1983, pp. 147-171.

Pencavel, John, An Analysis of the Quit Rate in American Manufacturing Industry, Industrial Relations Section, Princeton University, 1970.

Phelps, Edmund, "The Statistical Theory of Racism and Sexism," American Economic Review, September 1972, pp. 659-661.

Riche, Richard; Hecker, Daniel; and Burgan, John, "High Technology Today and Tomorrow: A Small Slice of the Employment Pie," Monthly Labor Review, November 1983, pp. 50-58.

Ryscavage, Paul, SIPP and CPS Labor Force Concepts. Paper presented at the American Statistical Association meeting, Philadelphia, Pa., August 1984.

Schiller, Bradley, Human Capital Transfers from Small to Large Businesses, U.S. Small Business Administration, Office of Advocacy, July 1982.

Spence, Michael, Market Signalling: Informational Transfer in Hiring and Related Screening Processes, Harvard Univeristy Press, 1974.

Theil, Henri, Economics and Information Theory, North-Holland Publishing Company, 1967.

Thurow, Lester, "The Disappearance of the Middle Class," The New York Times, February 5, 1984, p. 21.

Weiss, Leonard, "Concentration and Labor Earnings," American Economic Review, March 1966, pp. 96-117.

Welch, Finis, Minimum Wages: Issues and Evidence, American Enterprise Institute for Public Policy Research, Washington, D.C., 1978.