ELSEVIER

Journal of Econometrics 68 (1995) 153–179

JOURNAL OF Econometrics

How representative are matched cross-sections? Evidence from the Current Population Survey

Franco Peracchi*[a,b], Finis Welch[c]

[a]*Department of Economics, Università 'd'Annunzio', I-65100 Pescara, Italy*
[b]*IGIER, Abbazia di Mirasole, I-20090 Opera (MI), Italy*
[c]*Department of Economics, Texas A&M University, College Station, TX 77843, USA*

Abstract

In this paper we use the Current Population Survey (CPS) to illustrate some problems that arise in analyzing longitudinal data constructed by matching person records across files of rotating cross-section surveys. After studying the determinants of attrition from the CPS, we focus on two issues. The first is the effect of attrition on estimates of several labor market outcomes, such as participation rates, wages, and transitions rates between labor force states. The second is possible uses of the matched CPS to understand the nature of the measurement error process.

Key words: Attrition; CPS; Labor force participation; Matched cross-sections; Measurement errors; Rotating surveys; Wages
JEL classification: C42; J21; J31

1. Introduction

Longitudinal or panel data are routinely used to analyze individual labor force behavior. More recently, synthetic panels of group averages from repeated

* Corresponding author.

We thank Badi Baltagi and two anonymous referees for helpful comments. This research was funded by the AARP Andrus Foundation, the C.V. Starr Center for Applied Economics at New York University, and the U.S. Department of Health and Human Services, National Institute of Aging. Opinions and conclusions contained herein are solely those of the authors and should not be construed as representing opinions or policy of any agency of the Federal Government. The original data and tabulations of the data presented in graphical form in this paper are available from the authors upon request.

SSDI 0304407694016471

cross-sections have also been proposed for cases where genuine panels do not exist, or may be inferior to available cross-sections in terms of sample size, attrition problems and data quality (see, e.g., Deaton, 1985). In this paper, we sketch some of the issues that arise in using matched cross-sections, that is, longitudinal data constructed by matching person records across files of rotating cross-section surveys.

Our analysis is based on March–March matched files of the Current Population Survey (CPS). Some of the problems that we discuss are common to all panel data, but others are unique to this source. Because surveys in several countries adopt CPS-type sampling frames, however, the lesson learned from the matched CPS may have much broader applicability.

The first issue is attrition. The case of incomplete panel data has been studied, among others, by Fuller and Battese (1974), Bjorn (1981), Baltagi (1985), and Wansbeek and Kapteyn (1989). In all these studies, the individual decisions concerning participation in a panel survey are assumed to be independent of the behavioral relationship of interest. This assumption may be inappropriate in the case of matched cross-sections. The selection bias due to nonrandom attrition was first considered by Hausman and Wise (1979) in the context of estimation of linear models from panel data. More recently, Van den Berg, Lindebom, and Ridder (1991) model jointly the distribution of duration in a particular labor market state and the duration of panel survey participation. In this paper we study the determinants of attrition from the CPS, and investigate the type of biases that may arise by selecting those who can be matched.

The second issue is measurement error. Economists are well aware of the biases that may arise from poor measurement of explanatory variables. The measurement error problem may be particularly serious in the case of matched cross-sections if, as with the CPS, the agency that collects the data makes little attempt to check consistency of the responses across different surveys. Poterba and Summers (1986), for example, show that although measurement and reporting errors need not affect estimates of labor force participation, they may seriously affect estimates of turnover obtained from the month–month matched CPS. Economists, however, rarely ask the question: can measurement errors be predicted? To understand the measurement error process in survey data, detailed knowledge of the interview process is necessary. What is the form of the interview (face-to-face, telephone)? Who is the survey respondent (own response, response about others)? What is the potential impact of a particular sequence of questions leading to the response? Of considerable importance are also the data processing and cleaning procedures used in the post-interview phase of the survey. What responses are subject to clerical coding after the survey? What coding rules are followed? What are the criteria for imputation of missing values? These are crucial problems for empirical research, and economists should start addressing them directly rather than leaving them to others.

Although it is a little-explored and messy area, it is far easier to study than before thanks to modern computers, and may result in high pay-offs.

Somewhat paradoxically, the very reason that may lead to measurement error biases in the matched CPS, namely possible lack of consistency in responses across surveys, also makes these data an interesting source of information on the measurement error process for variables that are either time-invariant or change deterministically over time.

The remainder of this paper is organized as follows. Section 2 describes the data. Section 3 analyzes attrition due to matching. Section 4 leaves aside issues of measurement error and analyzes the effects of using the matched CPS to estimate participation and transition probabilities between labor force states, and aspects of the wage distribution. Section 5 leaves aside issues of attrition and discusses ways of using the matched CPS to understand the nature of the measurement error process. Section 6 contains some conclusions.

2. The data

The CPS is by far the most commonly used source of data on income and employment of the civilian noninstitutional population of the U.S.A. Administered by the Census Bureau, the CPS is a rotating monthly survey with rotation scheme 4–8–4, that is, a new rotation group enters the survey every month and each group is first interviewed for four consecutive months, temporarily dropped for eight months, and then re-interviewed for four consecutive months. Thus, each month sample consists of eight rotation groups labelled by the number of successive interviews, in any two-month period there are six overlapping rotation groups, and for any rotation group the second wave of four interviews occurs, one year later, in the same calendar months as the first wave.

Because of this sampling frame, if the sample size was constant and there was no attrition, three fourths of the sample in any two consecutive months could be matched. Further, in any given month, those in rotation groups 1–4 (the entering rotations) could be matched with those in rotation groups 5–8 (the outgoing rotations) the same month of the following year, while those in the outgoing rotations could be matched with those in the entering rotations the same month of the previous year. Matching person records across CPS files results in short panels that are used increasingly to study short-run labor force dynamics.

The advantages of using the CPS are well known: it is large – much larger than comparable alternatives, it has been available for public use for a long time, and is nationally representative. Many of the disadvantages are also well known. It is a 'rooftop' survey – there are no attempts to follow movers, for example. There is only one respondent per household, a 'knowledgeable' adult who answers for all residents. Nonresponse with imputations for missing values is

very common, and periodic changes in imputation methods have created interesting shifts in the data. The survey also changes periodically, and time-consistent measures are not always available. Public release CPS files contain no information on the form of the interview, and the respondent is identified only starting with 1990. Most importantly for our purposes, because the longitudinal aspect of the CPS is not deliberate but is rather a byproduct of the sampling frame, using the CPS as a panel is subject to additional problems and limitations that must be carefully considered by users of these data.

In this paper we disregard month-to-month dynamics, and focus on the matched CPS as a vehicle for analyzing dynamics over a one-year period. Matching is based on the March (Annual Demographic) files, which contain supplemental information on work experience and income in the previous calendar year.[1]

Each household in the CPS and each person within a household are assigned unique identifiers which are supposedly constant throughout the eight interviews. These identifiers are the primary vehicle for matching person records across adjacent years.[2] The 1979–91 March files used in this paper were matched using three steps of the algorithm described in Welch (1993).[3] Matching March CPS files generates a sequence of overlapping two-year panels with a large sample size. On average, each panel contains about 22,000 households and 57,000 persons (Tables 1 and 2). One important advantage of the matched CPS over existing panels, such as the National Longitudinal Survey or the Retirement History Survey, is that it spans several birth-cohorts and maintains a distribution by age and other person characteristics that is relatively stable, changing only with sample noise and with aggregate movements.

3. Attrition due to matching

Far from being random, success in matching person records across CPS files is systematically related to observable characteristics of both a person and his

[1] The CPS March files were first released for public use in 1968, the 1993 survey is currently the most recent.

[2] Household identifiers were scrambled in the 1972 survey, which therefore can be matched neither with the 1971 nor with the 1973 surveys. Identifiers were shifted in the 1977 and 1986 surveys, so 1977 cannot be matched with 1976 and 1986 cannot be matched with 1985. Thus, of the 24 pairs of adjacent-year March files between 1968 and 1992, only 20 can be matched. Person identifiers are unavailable from 1976 to 1978. For these years, only a person position number is available, which is not unique to a person but can change with changing family composition.

[3] The variables used at each step include the person identifier, sex, and race. Step one requires exact agreement on age, whereas the other two steps only require approximate agreement, that is, no age change or two-year changes are allowed. The tie-breaking part of each step is based on major activity last week, years of schooling completed, and relationship to the household head.

Table 1
Households in the CPS by match type

Year	Matched		Unmatched		Total
	Rot 1–4	Rot 5–8	Rot 1–4	Rot 5–8	
1979	20630	—	6816	—	27446
1980	25352	20630	7464	11791	65237
1981	22559	25352	9895	7924	65730
1982	22843	22559	6784	7089	59275
1983	22429	22843	7166	6771	59209
1984	21677	22429	7849	7207	59162
1985	—	21677	—	8398	30075
1986	21609	—	7849	—	29413
1987	21980	21609	7113	7570	58272
1988	20781	21908	8781	7332	58874
1989	21189	20781	6493	6872	55335
1990	22656	21189	7232	8864	59941
1991	—	22656	—	7599	30255

Table 2
Persons in the CPS by match type

Year	Matched		Unmatched				Total
			In matched hhs		In unmatched hhs		
	Rot 1–4	Rot 5–8	Rot 1–4	Rot 5–8	Rot 1–4	Rot 5–8	
1979	52631	—	6568	—	18277	—	77476
1980	64005	52631	7570	5707	19610	31965	181488
1981	56594	64005	6530	6577	26517	21135	181358
1982	56900	56594	6779	5802	18025	18603	162703
1983	55350	56900	7014	6357	18978	18036	162635
1984	53082	55350	6268	6169	21191	19092	161152
1985	—	53082	—	5450	—	22497	80939
1986	51315	—	6960	—	20160	—	78435
1987	52252	51315	7072	6278	18631	19920	155468
1988	48724	52252	6643	6277	22913	19097	155906
1989	49457	48724	6253	5881	16885	17487	144497
1990	53270	49457	7138	5802	18909	23503	158079
1991	—	53270	—	6503	—	20240	79317

household. After discussing the sources of match failure, we study separately the probability of matching households and the probability of matching persons in a matched household. A simple logit specification is used to model the dependence of match probabilities on a broad set of household and person attributes.

3.1. Sources of match failure

As with most household surveys, the basic CPS sample unit is an address not an actual household. In 1990 there were about 75,000 addresses assigned to be interviewed each month in the CPS. Because of vacancies and other reasons, on average only about 63,000 addresses contained persons eligible for interview, and of these only about 60,000 were actually interviewed. Occupants of a sample address who move away or die between two interviews are dropped, while any new occupant is included in the survey in the same rotation group as the other occupants. When all the previous occupants of a sample address move away, the new tenants are also included in the survey. They are assigned the same household identifier and the same rotation group as the previous occupants, but are flagged as new entrants. This sampling frame results in two sources of match failure. The first is failure to match any person in households that move in or out of the survey. The second is failure to match persons who move in or out of a matched household.

Besides migration, there are two other identifiable reasons why a household which is interviewed in one March survey has no interview records in the other March survey. One is noninterview, due either to refusal or inability to conclude an interview ('no one home'), or to construction or demolition that precludes occupancy of a dwelling place. A second is the practice by the Census Bureau of temporarily adding Hispanic households from the November survey to increase the size of the Hispanic subsample. Additional sources of match failure that cannot be separately identified using the public release March files are errors in transcribing the household identifier onto tape, and other addition or deletion of households by the Census Bureau for purposes of sample design.[4]

Fig. 1 presents summary statistics for the entering rotations of each matched CPS from 1979 to 1990. The top-left panel presents the overall match rates. On average, two thirds of the people can successfully be matched. Match rates, however, vary considerably over both rotation group and time. They are systematically lowest for the first and highest for the fourth rotation group, which is consistent with the evidence from other studies of a declining hazard of exit from a panel survey. Over time, they drop substantially in 1981 and 1988 due to deletion of households from the CPS.[5]

[4] Pitts (1988) analyzes the distribution of unmatched households by reason of match failure. He finds that, on average over the period 1979–83, 42 percent are 'movers', 29 percent are noninterviews, 12 percent are oversampled Hispanics, while for the other 17 percent the reason cannot be discerned using the public release tapes. Interestingly, only one-third of the noninterviews are nonresponses or 'no one home'.

[5] The drop for the first rotation group in 1984 is due instead to a coding error in the subfamily identifier for a set of about 700 households.

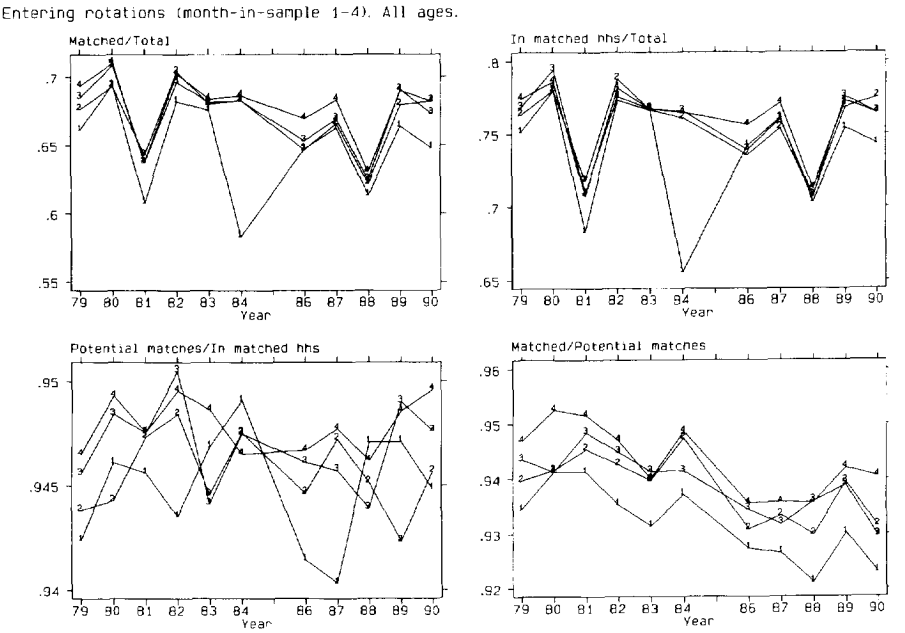

Fig. 1. Match statistics.

The other three panels of Fig. 1 provide a factorization of the overall match rate. The top-right panel shows the fraction of people in matched households. The bottom-left panel shows the ratio of potential matches to the number of people in matched households, with the number of potential matches defined as the smallest size of a matched household in the two surveys. The bottom-right panel shows the ratio of the number of people actually matched to the number of potential matches. This ratio may be taken as an indicator of how conservative is our matching algorithm. On average, three fourths of the unmatched are in unmatched households. Potential matches are remarkably stable at 94.5 percent of the number of people in matched households, whereas actual matches show a slight downward trend from 94 and to 93 percent of the potential matches. These data indicate that while one cannot expect major gains by improving our algorithm for matching persons across CPS surveys, substantial gains could instead be obtained from an effort by the Census Bureau to increase household match rates, in particular by following 'mover' households.

In Fig. 2 we present overall match failure rates by age and sex, separately for the entering and the outgoing rotations, along with their two components, the fraction of people in unmatched households and the fraction of unmatched people in matched households. We exclude newborns (age 0) because they obviously have no previous year's record.

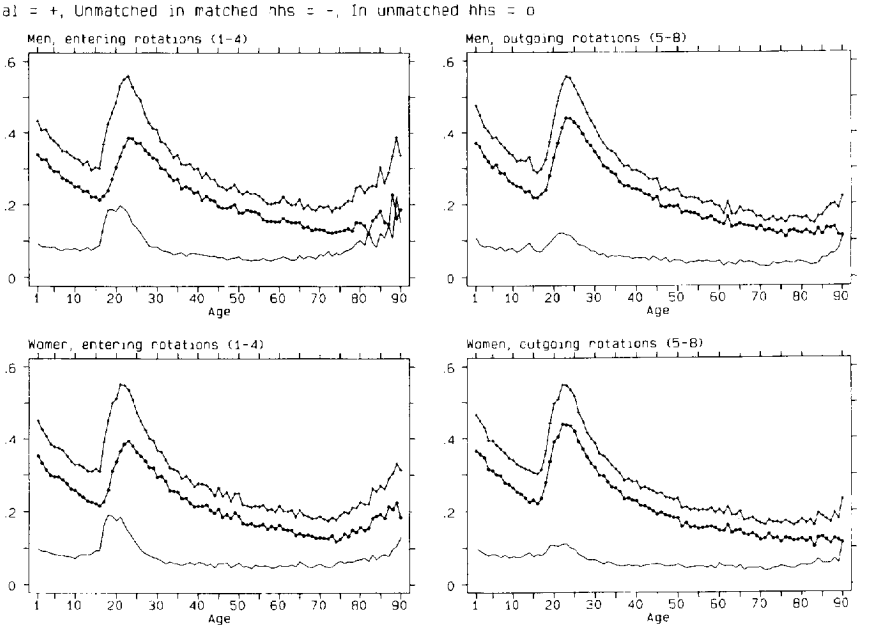

Fig. 2. Match failure rates.

The age profiles of the two components are quite different. The age profiles of the fraction of people in unmatched households do not differ much by sex and rotation group. They start off at about 45 percent, decline up to age 16, then rise sharply, reaching a peak of about 55 percent at age 23 for men and age 21 for women. They drop steadily afterwards, reaching a minimum of about 18 percent around age 70, when they start rising again, especially for the entering rotations. Assuming a 20–25 average age differentials between parents and their children, this pattern is consistent with young parents moving along with their infant children.

On the other hand, the age profiles of the fraction of unmatched people in matched households are similar for men and women, but differ by rotation group. For the outgoing rotations they are fairly stable at about 10 percent, whereas for the entering rotations they rise dramatically between age 17 and age 22. While on average only one unmatched out of four is in matched households, this fraction rises to two out of five for people aged 17 to 22 in the entering rotations.

Thus, excluding newborns, the main source of attrition in the CPS is failure to follow people of college age in matched households and young households who move. Overall, college enrollment and young households mobility account for about one-third of the match failures for the entering rotations, while

newborns and young households mobility account for about one-third of the match failures for the outgoing rotations. Our results also suggest that failure to match people in matched households occurs for different reasons than failure to match households. The two processes ought to be considered separately.

3.2. *Household match probabilities*

In this section we study the relationship between household match probabilities and observable household characteristics. We do not distinguish, however, between reasons for household match failure, such as residential mobility, nonresponse, or sample redesign by the Census Bureau. Partly this is due to our desire to keep the analysis as simple as possible, and partly to data limitations. Mobility data from the 1979 and 1980 CPS take March 1975 as reference, those from the 1985 CPS refer to March 1980, while only for the remaining surveys the reference is to the previous year's March.

The first two columns of Table 3 present the results of fitting a linear logit model to household data from 1979 to 1991, separately for the entering and the outgoing rotations. The response variable is a binary indicator for success in matching a household. The predictors include characteristics of both the household (household size, presence of children, region of residence, rotation group, and survey year) and the head (age, sex, race, schooling level,[6] Armed Forces enlistment, marital status, and health status). The introduction of year and rotation group dummies should help control for the effects of sample redesign by the Census Bureau. We restrict attention to households whose head is aged 22–74 years at the time of the survey. Age of the head enters as a set of dummies for roughly five-year age intervals. The baseline is a two-person household resident in the North-East, in the first rotation group in 1980, headed by a white man aged 30–34, high-school graduate, not in the Armed Forces, not ill, married with spouse present. Baseline probabilities are reported at the bottom of the page, along with the sample size and standard measures of goodness of fit.

Household match probabilities increase monotonically with the age of the head. If retirement from the labor force of the household head was associated with geographical mobility, we would expect match rates to drop for households whose head is 62 or older. We see no evidence of this. Controlling for all other characteristics, we find that a female headed-household is more likely to be matched than a male-headed one. A household headed by a black is more likely to be matched than one headed by a white, while the opposite is true if the head is another nonwhite. A household is less likely to be matched if the head is in the Armed Forces, or did not finish high-school, or is unmarried or ill. Match

[6] We define four schooling levels: High-school dropouts (less than 12 years of schooling), high-school graduates (12 years), college dropouts (between 13 and 15 years), and college graduates (16 years or more).

Table 3
Match probabilities: Logit coefficients (* denotes significance at the 10% level, ** denotes significance at the 5% level)

	Unmatched households		Unmatched persons	
Variables	Rot 1–4	Rot 5–8	Rot 1–4	Rot 5–8
Constant	1.247**	1.062**	2.685**	2.904**
Female	0.097**	0.144**	0.017*	0.083**
Age 1–16	—	—	0.947**	0.425**
Age 17–21	—	—	−0.217**	0.053**
Age 22–24	−0.769**	−1.109**	−0.643**	−0.620**
Age 25–29	−0.340**	−0.456**	−0.304**	−0.331**
Age 35–39	0.244**	0.282**	0.105**	0.192**
Age 40–44	0.423**	0.464**	0.162**	0.256**
Age 45–49	0.593**	0.700**	0.274**	0.334**
Age 50–54	0.771**	0.892**	0.362**	0.464**
Age 55–61	0.961**	1.121**	0.422**	0.516**
Age 62–64	1.083**	1.252**	0.459**	0.622**
Age 65–69	1.250**	1.349**	0.405**	0.663**
Age 70–74	1.424**	1.546**	0.418**	0.888**
Age 75–79	—	—	0.315**	0.824**
Black	0.143**	0.143**	−0.079**	−0.140**
Other nonwh.	−0.042	−0.124**	−0.192**	−0.251**
HS dropout	−0.340**	−0.337**	−0.183**	−0.237**
College drop.	−0.058**	−0.040**	0.061**	0.149**
College grad.	0.041**	−0.022*	0.072**	0.088**
In the AF	−0.715**	−0.598**	−0.514**	−0.661**
Ill	−0.139**	−0.115**	−0.216**	−0.091**
Ever married	−0.595**	−0.575**	−0.198**	−0.551**
Never married	−0.427**	−0.376**	0.150**	0.313**
Family size	0.021**	0.016**	—	—
Hh with kids	0.064**	0.134**	—	—
Spouse	—	—	0.021	−0.330**
Child	—	—	−1.412**	−1.220**
Nonrelative	—	—	−2.005**	−2.491**
South	0.050**	0.102**	0.134**	0.242**
Midwest	−0.159**	−0.174**	−0.151**	−0.123**
West	−0.486**	−0.499**	−0.338**	−0.344**
Rot 2	0.095**	—	0.059**	—
Rot 3	0.104**	—	0.077**	—
Rot 4	0.122**	—	0.130**	—
Rot 6	—	0.167**	—	0.072**
Rot 7	—	0.177**	—	0.098**
Rot 8	—	0.167**	—	0.150**
Year 79	−0.160**	—	−0.069**	—
Year 80	—	−0.688**	—	−0.070**
Year 81	−0.450**	—	0.033*	—
Year 82	−0.043**	−0.025	0.002	0.008
Year 83	−0.125**	0.007**	−0.052**	−0.095**
Year 84	−0.279**	−0.075**	0.011	−0.089**
Year 85	—	−0.305**	—	−0.004
Year 86	−0.304**	—	−0.161**	—
Year 87	−0.190**	−0.219**	−0.160**	−0.226**
Year 88	−0.492**	−0.174**	−0.179**	−0.217**
Year 89	−0.143**	−0.179**	−0.124**	−0.205**
Year 90	−0.182**	−0.448**	−0.152**	−0.229**
Year 91	—	−0.207**	—	−0.206**
Baseline probab.	0.777	0.743	0.936	0.948
Obs.	291973	294987	652950	643217
$\ln \hat{L}_n^0$	−165167.8	−170809.7	−228534.1	−212032.8
$\ln \hat{L}_n$	−152984.5	−154476.7	−209390.6	−194782.6
Pseudo R^2	0.0738	0.0956	0.0838	0.0814

probabilities increase with household size, and are higher if a household contains children less than 18 years of age. They differ considerably by geographical region: they are higher in the South, and lower in the Midwest and especially the West. Finally, they increase monotonically with duration in the survey.

3.3. Match probabilities within matched households

We now turn to the probability of matching people within matched households. Our basic model is again a linear logit model, where the response variable is now a binary indicator for success in matching a person in a matched household. The model was fitted to individual data from 1979 to 1991, with the sample restricted to persons in matched households, aged 1–79 at the time of the survey.

The predictors include dummies for age, sex, race, schooling level, Armed Forces enlistment, marital status, relationship to the household head, health status, geographical region, rotation group, and survey year. The baseline is a white man aged 30 to 34, household head, high-school graduate, married with spouse present, living in the North-East, not in the Armed Forces, not ill, in the first rotation group in 1980.

The last two columns of Table 3 present the estimated logit coefficients, separately for the entering and the outgoing rotations. The implied age profiles of match probabilities agree nicely with the ones in Fig. 2.

Controlling for all other characteristics, we find that match probabilities are a little higher for women than for men. They also increase with schooling level and duration in the survey. Nonwhites, those in the Armed Forces, those who are ill, or those who reside in the Midwest or the West are less likely to be matched. Marital status and relationship to the household head are also important. Match probabilities are higher for those who never married, and lower for those who are divorced or separated. For the entering rotations, spouses have about the same match probabilities as household heads. Spouses in the outgoing rotations, however, are less likely to be matched, presumably because of new arrival into sample households through marriage or permanent cohabitation. Finally, match probabilities are substantially lower for children, other relatives of the head, and other household members.

4. Attrition biases

In this section we ask whether labor market outcomes are independent of the process that determines attrition. Unless this is true, the sample information provided by the matched CPS may not identify population characteristics of interest. Our approach is based on comparing the information contained in the matched and the unmatched CPS. Systematic differences between the two sources will be evidence against the hypothesis of independence, and will be

indicative of the kind of biases that may arise by selecting those who can be matched. What bias may arise by using the matched CPS clearly depends on the population feature of interest. In the next three sections we focus on the joint distribution of labor force status and wages at a given survey, and the joint distribution by labor force status in two adjacent surveys.

Let W and Y denote the response variables, respectively wages and an indicator of labor force status at a given survey, let X be a vector of predictors, and let D be an indicator for the type of match, taking value 1 for the matched, 2 for the unmatched in matched households, and 3 for those in unmatched households. The population joint distribution of labor force status and wages at a given survey can be decomposed as

$$F_x(W, Y) = \sum_{d=1}^{3} F_x(W \mid Y, D = d) P_x(Y \mid D = d) P_x(D = d), \tag{1}$$

where P_x denotes the conditional probability and F_x the conditional distribution function given $X = x$. The match probabilities $P_x(D = d)$ have already been analyzed in Section 3. In Sections 4.1 and 4.2 we study whether $P_x(Y \mid D = d)$ and $F_x(W \mid Y, D = d)$ differ systematically by match type and possibly by rotation group.

Similarly, the population joint distribution by labor force status in two adjacent surveys can be decomposed as

$$P_x(Y, Y') = \sum_{d=1}^{3} P_x(Y' \mid Y, D = d) P_x(Y \mid D = d) P_x(D = d). \tag{2}$$

In the absence of prior information, the one-year transition probabilities $F_x(Y' \mid Y, D = d)$ are only identifiable for the matched ($D = 1$). Thus, the population distribution (2) may not be identifiable from the matched files. A strong identifying assumption is that labor force transitions are conditionally independent of D given X. Under this assumption, transition probabilities of the matched are the same as those of the unmatched. In Section 4.3 we study whether this independence assumption is supported by the data.

Notice that the choice of predictors in X is important, for it may well be that observed differences between the matched and the unmatched, and between different rotation groups, disappear after controlling for a sufficiently broad set of person characteristics.

4.1. Participation probabilities

We first ask whether participation behavior differs systematically by match type and rotation group. Following Peracchi and Welch (1994), we classify people into three mutually exclusive labor force states, namely full-time work, part-time work, and out of the labor force. Our classification is a 'spot' one,

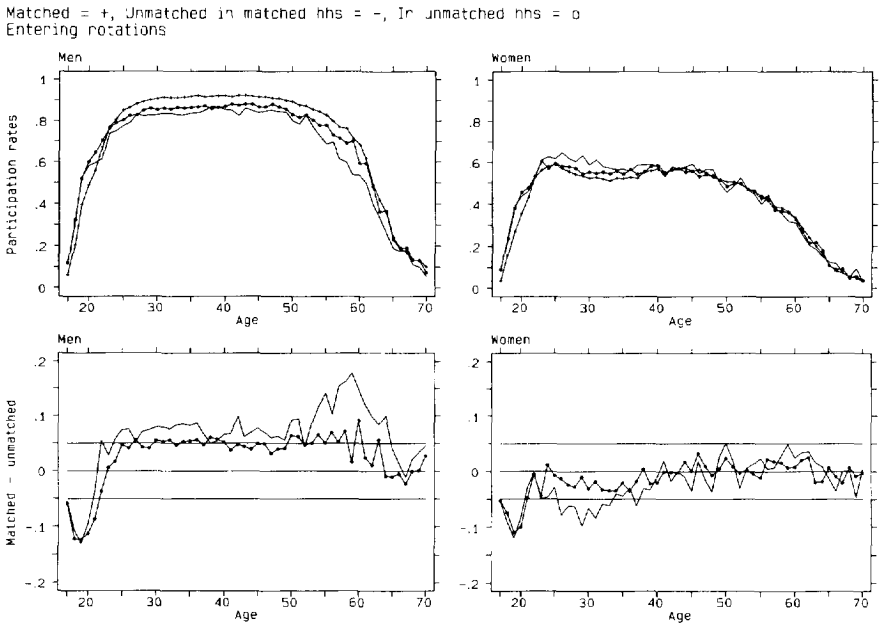

Fig. 3. Full-time age–participation profiles.

based mainly on hours worked in the week preceding the survey.[7] For the younger people, being out of the labor force is closely associated with schooling. For the older people, it is closely associated with retirement.

In Fig. 3 we compare full-time age–participation profiles of men and women aged 17–70 for the matched and the unmatched. For reasons of space, only the results for the entering rotations are presented. The data are simple averages of full-time participation frequencies for cells defined by sex, age, and survey year.

Consider men first. Except at young ages (17–22), matched men have higher participation rates than the unmatched. Among the unmatched, those in unmatched households always have higher participation rates than the unmatched in matched households. For the former, the difference with respect to the

[7] Specifically, a person is a full-time worker if he/she had a job in the week preceding the survey and worked at least 35 hours or, if worked less than 35 hours, usually works 35 hours per week or has a full-time work week of less than 35 hours. Those unemployed are classified as full-time worker if they are looking for a full-time job, have been unemployed for not more than one year, and worked at least two weeks sometimes in the previous five years.

Those who have a job are classified as part-time if they are not full-time. Similarly, an unemployed person is classified as part-time if he/she is not full-time. Finally, one who is neither full-time nor part-time is classified as out of the labor force.

matched is stable at about 5 percentage points. For the latter, it is nearly 10 percentage points, and grows to about 15 percentage points for those aged 55–60.

For women, differences in participation by match type are smaller and almost disappear after age 40. Even at younger ages, however, the patterns are different from those of men: Unmatched women have higher participation rates than the matched, while unmatched women in matched households have higher participation rates than women in unmatched households.

How much of the differences between matched types remains after controlling for other person characteristics? In what dimensions does the conditional distribution by labor force status given observable person characteristics differ by match types? To answer these questions, we estimate multinomial logit (MNL) models for participation by sex and match type. The response variable is an indicator of labor force status. The baseline is a white, high-school graduate, aged 30–34, household head, married with spouse present, with no children living in the household, not ill, resident in the North-East, in the first rotation group in 1980.

In Table 4 we present the estimated MNL coefficients for the odds of being out of the labor force relative to full-time work. The first two columns of the table present the point estimates for the matched, separately for the entering and the outgoing rotations. The other four columns report the estimated differences between the matched and the unmatched. At the bottom of the table we report the baseline probabilities of being in each labor force state for the matched, along with the differences with respect to the unmatched.

Although the differences between match types are somewhat smaller than in the case of Fig. 3, differences in the coefficients of key variables such as age, race, schooling level, and relationship to the household head are often statistically significant. For example, the odds of being out of the labor force are always higher for blacks and for those who did not complete high-school. Interestingly, however, the race effect is much stronger among the unmatched, while the second effect is significantly attenuated.

4.2. Wages

Fig. 4 compares the age profiles of a measure of center – the median – and a measure of dispersion – the interquartile range – of full-time weekly wages for the three match types. Our wage measure is real annual earnings (in 1982$) in the year preceding the March survey divided by weeks worked in that year. The definition of full-time is slightly different from the one adopted in Sections 4.1 and 4.3. Full-time workers are here those who worked at least 50 weeks in the calendar year before the March survey, usually at least 35 hours per week ('full-time year-round'). Self-employed and farm workers are excluded from the sample. The data are simple averages of medians and interquartile ranges for

Table 4
Participation probabilities: MNL coefficients for out of the labor force – Men (* denotes significance at the 10% level, **denotes significance at the 5% level)

			Differences w.r.t. unmatched			
	Matched		In matched hhs		In unmatched hhs	
Variables	Rot 1–4	Rot 5–8	Rot 1–4	Rot 5–8	Rot 1–4	Rot 5–8
Constant	− 3.962**	− 3.864**	− 0.609**	− 0.573**	− 0.792**	− 0.727**
Age 17–21	1.958**	2.095**	0.663**	0.562**	0.595**	0.671**
Age 22–24	0.475**	0.617**	0.307**	0.155	0.053	0.126*
Age 25–29	0.046	0.139**	0.170*	0.042	− 0.107*	− 0.039
Age 35–39	− 0.000	0.056	0.042	− 0.232**	− 0.050	0.048
Age 40–44	− 0.029	0.053	− 0.010	− 0.141	0.114	0.052
Age 45–49	0.082*	0.125**	− 0.042	− 0.330**	0.146*	0.043
Age 50–54	0.487**	0.521**	− 0.060	− 0.299**	0.113	0.122
Age 55–61	1.489**	1.460**	− 0.055	0.050	0.317**	0.154**
Age 62–64	3.130**	3.172**	0.210	0.127	0.540**	0.270**
Age 65–69	4.530**	4.601**	0.452**	0.542**	0.679**	0.475**
Black	0.366**	0.477**	− 0.229**	− 0.056	− 0.288**	− 0.153**
Other nonwh.	0.430**	0.387**	− 0.249**	− 0.221**	− 0.389**	− 0.363**
HS dropout	0.525**	0.590**	0.066	0.222**	0.287**	0.324**
College drop.	0.345**	0.354**	− 0.405**	− 0.137**	− 0.029	− 0.029
College grad.	− 0.254**	− 0.287**	− 0.409**	− 0.344**	− 0.171**	− 0.266**
Ill	2.590**	2.612**	0.419**	0.373**	0.435**	0.404**
Ever married	0.330**	0.364**	0.033	0.078	0.178**	0.300**
Never married	0.622**	0.687**	− 0.090	0.012	0.291**	0.472**
Spouse	0.224**	0.220**	− 0.027	− 0.006	− 0.047	0.072
Relative	1.008**	0.838**	0.459**	0.258**	0.139**	− 0.114**
Other	0.223**	0.090	0.127	− 0.036	− 0.006	− 0.220**
Hh with kids	0.036*	0.034*	− 0.128**	− 0.128**	− 0.137**	− 0.117**
South	0.077**	0.075**	0.130**	0.215**	− 0.087**	0.069
Midwest	0.203**	0.140**	0.358**	0.223**	0.112**	0.101**
West	0.327**	0.310**	0.324**	0.288**	0.084**	0.118**
Baseline probab.						
Full-time	0.972	0.969	0.018	0.024	0.026	0.023
Part-time	0.010	0.011	− 0.003	− 0.009	− 0.006	− 0.002
Out of labor force	0.018	0.020	− 0.015	− 0.015	− 0.020	− 0.021

cells defined by sex, age, and survey year. To reduce the sampling noise, age profiles have been smoothed using Cleveland's (1979) locally linear scatterplot smoother (loess). For reasons of space, we only present the results for the entering rotations.

For women, the differences between match types are very small. For men, we observe systematic differences in median wages, but little differences in wage dispersion. Except at younger ages, the pattern of median male wages resembles

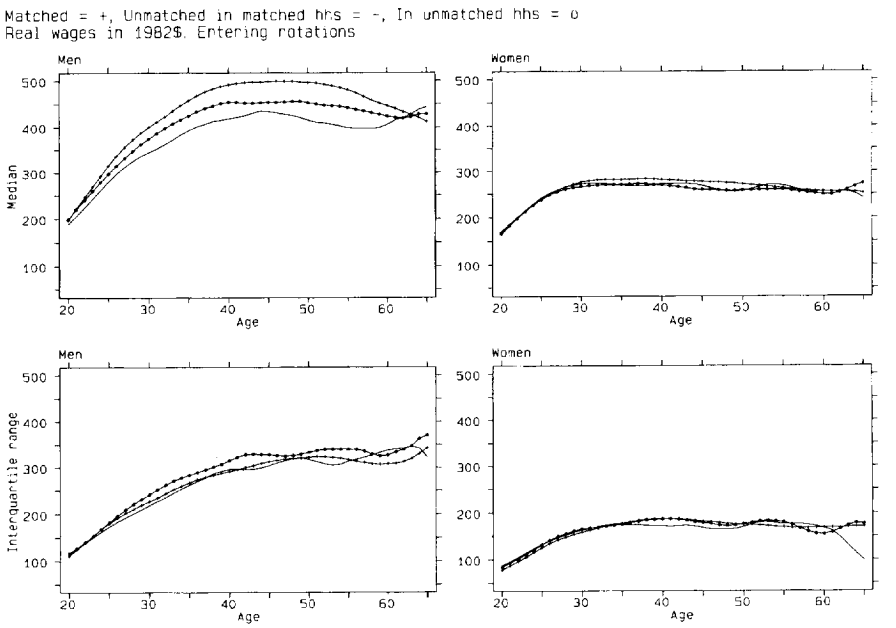

Fig. 4. Full-time year-round weekly wages by age.

the one of participation rates: Median wages are always highest for the matched and lowest for the unmatched in matched households.

To get a better understanding of the differences between match types, we also fitted linear models to the conditional mean, the conditional median, and the conditional upper and lower quartiles of male wages, separately for the entering and the outgoing rotations. The sample was restricted to full-time year-round workers with annual earnings above 100$ and with 1 to 40 years of labor force experience.[8] Self-employed, farm workers, and other nonwhites were also excluded. The response variable is the natural logarithm of weekly wages. Following Murphy and Welch (1992), each model is specified as quadratic in the transformed experience variable $z = (\text{experience} - 30)^2/100$. The predictors also include a set of dummies for race, schooling level, and calendar year.

Table 5 presents the estimated coefficients, along with estimates of the average slope of the wage–experience profile for new entrants (1–10 years of experience) and peak earners (26–30 years of experience). Measures of statistical significance,

[8] Years of labor force experience are defined as age − 16 for high-school dropouts, age − 18 for high-school graduates, age − 20 for college dropouts, and age − 22 for college graduates.

based on heteroskedasticity-consistent estimates of precision, are only reported for the mean regressions, estimated by ordinary least squares (LS).[9]

The matched show higher mean and median wages, but now also larger wage dispersion, and a wider black/white differential. On the other hand, the unmatched show a wider differential between high-school graduates and high-school dropouts but no appreciable difference in the college premium with respect to the matched, and lower returns to experience for the new entrants.

4.3. *Transition probabilities*

We now ask what can be learned from the matched CPS about the probability that a person who is in labor force state i in the first survey will be in state j one year later. Direct comparison of transition rates between the matched and the unmatched is impossible in this case, for the unmatched CPS provides no direct information on transitions. We shall approach this problem in two ways. The first, due to Manski (1989), provides a quantitative assessment of the uncertainty about population transition probabilities after observing the selected data. The second uses grouped data from the matched and the unmatched CPS to derive indirect information about equality of population transitions for the different match types.

Let Y_{ij} take value one if a transition from i to j is observed and value zero otherwise, and let X be a random vector consisting of person characteristics. The objective is to learn about the discrete (one-year) transition probability $\lambda_{ij}(x) = P_x(Y_{ij} = 1)$. The selection problem is attrition from the sample. Let D_i be the indicator for the type of match for people in labor force state i in the first survey (as before, $D_i = 1$ for the matched). Since Y_{ij} is a binary random variable, $\lambda_{ij}(x)$ must lie in the closed interval

$$[E_x Y_{ij} 1(D_i = 1),\ E_x Y_{ij} 1(D_i = 1) + P_x(D_i \neq 1)] \tag{3}$$

(see Manski, 1989), where $1(\cdot)$ denotes the indicator function of an event and E_x denotes conditional expectations given $X = x$. Clearly, transition probabilities for the matched must also lie in (3). The lower bound is the transition probability if all the unmatched remain in state i, while the upper bound is the transition probability if they all move to state j. The width of the interval is equal to the conditional probability of match failure and may be viewed as a measure of how 'vague' is the information about $\lambda_{ij}(x)$ contained in the matched CPS. The higher the match rate, the tighter the bound.

[9] It is far from obvious how to construct heteroskedasticity-consistent estimates of precision for quantile regressions.

Table 5
Coefficients on wage regressions – Men (* denotes significance at the 10% level, ** denotes significance at the 5% level)

			Difference w.r.t. unmatched			
	Matched		In matched hhs		In unmatched hhs	
Variables	Rot 1–4	Rot 5–8	Rot 1–4	Rot 5–8	Rot 1–4	Rot 5–8
Mean						
Constant	5.898**	5.976**	0.106**	0.114**	0.064**	0.066**
Black	−0.227**	−0.229**	−0.040*	−0.066**	−0.024*	−0.038**
HS dropout	−0.261**	−0.263**	0.062**	0.065**	0.050**	0.067**
College drop.	0.153**	0.148**	0.005	0.012	0.019*	0.001
College grad.	0.401**	0.407**	−0.000	0.006	−0.015	−0.012
z	−0.045**	−0.041**	0.002	0.015	−0.004	−0.003
z^2	−0.007**	−0.007**	−0.000	−0.004**	−0.000	−0.001**
New entrant	0.061	0.063	0.004**	0.006*	0.017*	0.010*
Peak earner	0.002**	0.002**	−0.000	−0.000	−0.000	−0.000
Lower quartile						
Constant	5.665	5.742	0.148	0.121	0.108	0.109
Black	−0.244	−0.249	−0.052	−0.077	−0.035	−0.067
HS dropout	−0.307	−0.306	0.046	0.054	0.017	0.063
College drop.	0.156	0.147	0.003	0.013	0.010	−0.009
College grad.	0.356	0.363	−0.024	−0.026	−0.046	−0.045
z	−0.041	−0.033	−0.016	0.010	−0.008	−0.001
z^2	−0.007	−0.008	0.001	−0.003	−0.000	−0.002
New entrant	0.060	0.065	0.000	0.009	0.014	0.015
Peak earner	0.002	0.002	−0.000	0.000	−0.000	0.000
Median						
Constant	5.953	6.026	0.098	0.120	0.065	0.054
Black	−0.232	−0.233	−0.019	−0.055	−0.031	−0.033
HS dropout	−0.264	−0.271	0.068	0.082	0.071	0.089
College drop.	0.145	0.140	−0.013	−0.012	0.000	−0.011
College grad.	0.372	0.378	−0.048	−0.057	−0.031	−0.029
z	−0.048	−0.044	0.001	0.013	−0.006	0.000
z^2	−0.006	−0.007	−0.000	−0.004	−0.000	−0.002
New entrant	0.059	0.063	0.003	0.0003	0.015	0.011
Peak earner	0.002	0.002	−0.000	0.000	−0.000	0.000
Upper quartile						
Constant	6.180	6.248	0.035	0.062	−0.008	−0.012
Black	−0.212	−0.216	−0.044	−0.033	0.004	−0.037
HS dropout	−0.217	−0.219	0.094	0.093	0.082	0.088
College drop.	0.144	0.145	−0.004	−0.025	0.018	0.001
College grad.	0.429	0.443	−0.013	0.009	0.005	0.009
z	−0.049	−0.047	0.017	0.020	0.000	0.007
z^2	−0.006	−0.006	−0.002	−0.004	−0.000	0.002
New entrant	0.059	0.062	0.006	0.005	0.015	0.011
Peak earner	0.002	0.002	−0.000	−0.000	−0.000	−0.000

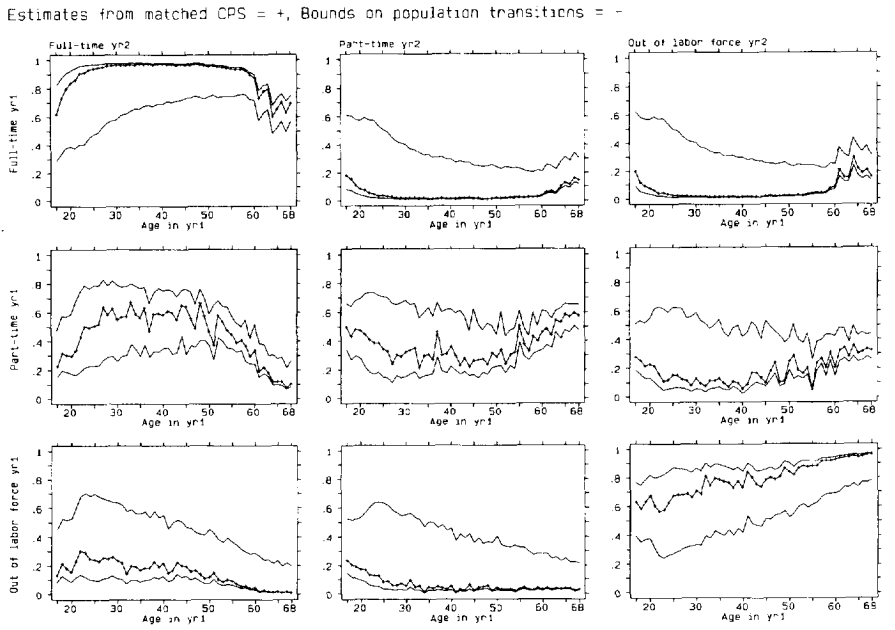

Fig. 5. Transition rates by age – Men.

If all variables in X are discrete, the interval (3) can be estimated consistently by

$$[b_{ij}(x), b_{ij}(x) + c_i(x)],$$

where

$$b_{ij}(x) = \frac{\#\{\text{state } j \text{ in year } t+1 | X = x\}}{\#\{\text{state } i \text{ in year } t | X = x\}},$$

$$c_i(x) = \frac{\#\{\text{unmatched} | X = x\}}{\#\{\text{state } i \text{ in year } t | X = x\}}.$$

In Figs. 5 and 6 we present the estimated bounds, along with estimates of the transition probabilities for the matched, separately by sex, age, and labor force status in the first year. The data are simple averages of observed frequencies for groups defined by sex, age, year, and labor force status in the first year. The substantial amount of noise in male transitions from part-time is due to the small fraction of part-time working men.

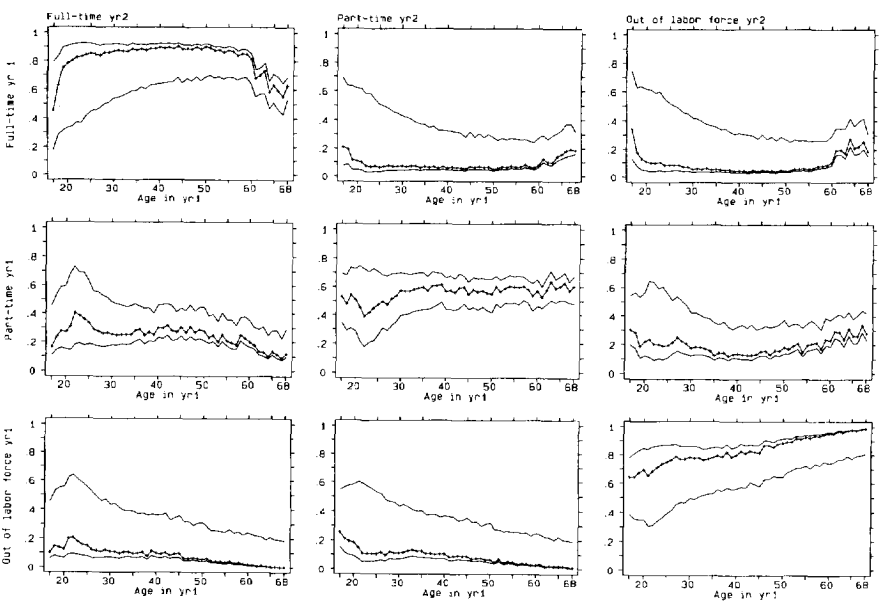

Fig. 6. Transition rates by age – Women.

Match failure rates – the width of the bound – differ by labor force status in the first survey. They seem to be higher for those who are full-time workers and lowest for those who are out of the labor force. The bounds are quite large, but they become tighter for people of older ages. The lower bounds on exit from full-time and from out of the labor force provide conservative estimates for the probabilities of exit. Transition rates for the matched lie in the bound, as they should. It is worth stressing that their position is uninformative about the position of population transition rates. In particular, the fact that, with the exception of exit from part time, observed exit rates for the matched are very close to the lower bound, simply reflects the fact that exit rates for the matched are small. Notice that, after controlling for labor force status in the first year, men and women do not differ much in terms of labor force behavior. The main difference is the role of part-time. Not only more women are employed in part-time than men, but average duration in part-time is longer for women than for men. Further, after age 30, there is little evidence of age dependence in female duration in part-time.

The second approach starts from the basic decomposition

$$\lambda_{ij}(x) = \sum_{d} E_x(Y_{ij}|D_i = d)\, P_x(D_i = d). \tag{4}$$

If $D_i = 1$ denotes a match and $P_x(D_1 = 1) > 0$, the identity (4) may be rewritten as

$$E_x(Y_{ij}|D_i = 1) = \lambda_{ij}(x)\frac{1}{P_x(D_i = 1)} - \sum_{d>1} E_x(Y_{ij}|D_i = d)\frac{P_x(D_i = d)}{P_x(D_i = 1)}. \quad (5)$$

We propose to estimate the regression model (5) using grouped data and then test for selectivity bias by looking at whether the estimated coefficient on $1/P_x(D_i = 1)$ is significantly different from (minus) the coefficients on the match odds ratios $P_x(D_i = d)/P_x(D_i = 1)$. Of course, only sample match odds ratios can be used as regressors. Although these are error-ridden measurement of their population counterparts, the resulting measurement-error bias is likely to be small if the sample size in each group is sufficiently large that sampling variability can be ignored.

In Table 6 we present the achieved significance levels (p-values) of F-tests for equality of the coefficients in the regression model (5). Data were grouped by sex, age, and U.S. state. For each sex–age combination, the model was estimated by weighted LS, with weights equal to the sample size in each age–sex–state cell. For reasons of space, we only present the results for exit from full-time and ages 21–60.[10] Significance levels exceeding the conventional 10 percent are simply denoted by '—'.

For men, out of 40 age-specific regressions, equality of regression coefficients is rejected at the 5 percent level seven times for the full-time to full-time case, six times for the full-time to part-time case, and five times for the full-time to out of the labor force case. Things look even better for women, where rejection at the 5 percent level occurs three times for the full-time to full-time and full-time to out of the labor force case, and two times for the full-time to part-time case. We conclude that, although selecting the matched individuals does bias measures of participation, especially for men, no systematic bias appears in the estimates of transitions after controlling for sex, age, and labor force status at the time of the first survey.

5. Measurement errors

Measurement errors may arise at various phases of a survey. Some of these errors are purely unpredictable, as assumed by the classical measurement error, but others are not, and depend systematically on the nature of the variables and the phase of the survey.

[10] Cell size vary considerably with sex and age. The largest median age–state group size for full-time workers is 82 observations for men aged 29 and 57 observations for women aged 28. The smallest median group size is 34 observations for men aged 59 and 19 observations for women aged 60.

Table 6
Significance probabilities (p-values) of F-tests for equality of the coefficients in the regression model (5)

	Men			Women		
Year 1 age	FT to FT	FT to PT	FT to RT	FT to FT	FT to PT	FT to RT
21	—	—	—	0.011	0.053	—
22	0.022	—	0.010	0.098	0.010	—
23	—	—	—	—	—	—
24	—	0.072	—	—	—	—
25	—	0.051	—	—	—	—
26	0.003	—	0.033	0.006	0.032	—
27	0.057	0.003	—	—	—	—
28	—	0.082	—	—	—	—
29	—	—	—	0.054	—	0.000
30	—	—	—	—	—	0.089
31	—	0.099	0.022	—	—	—
32	—	—	—	—	—	—
33	—	0.010	—	—	0.095	—
34	0.014	0.082	—	—	—	0.020
35	0.076	—	—	—	—	—
36	0.059	—	0.095	—	—	—
37	—	0.013	—	—	—	—
38	—	—	—	—	—	—
39	—	—	—	—	—	—
40	0.010	0.038	—	—	—	—
41	—	—	0.043	—	—	—
42	—	—	—	—	—	—
43	0.010	0.029	—	—	0.072	0.057
44	—	—	—	—	—	—
45	—	—	—	—	—	—
46	—	—	0.065	—	—	—
47	—	—	—	—	—	—
48	—	—	0.066	—	—	0.071
49	0.052	—	0.068	0.042	—	0.026
50	—	—	—	—	—	—
51	0.041	0.052	—	—	—	—
52	—	—	—	—	—	—
53	—	—	0.062	—	—	0.081
54	—	0.015	—	—	—	0.070
55	—	—	—	—	—	—
56	—	—	—	—	—	—
57	0.030	—	0.009	—	—	—
58	—	—	—	—	—	—
59	—	—	—	—	—	—
60	—	—	—	—	—	—

There are several strategies for analyzing measurement errors in survey data. In the absence of strong prior information on the error process, each requires repeated measurements of the same quantity. One strategy is to use re-interview data collected at short distance from the original interview. This strategy was followed for example by Poterba and Summers (1986), who used the CPS Reinterview Survey to analyze measurement errors in labor force status. If re-interview data are not available or are of dubious quality, a second strategy is to use matched cross-sections.

Consider the matched CPS. Its usefulness for analyzing measurement errors comes from the fact that the Census Bureau makes no attempt to check consistency of person records across survey. Thus, matched files provide repeated measurements on the same variable. Clearly, only errors in variables that are either time-invariant or evolve deterministically over time can be studied this way. Whether these repeated measurements may be treated as independent depends on various considerations. For example, in the case of industry and occupation of employment, which are subject to post-survey coding, the independence assumption may not be unrealistic. In other cases, for example when the respondent is the same in both surveys, the independence assumption may no longer be valid.

In the remainder of this section, we try to give a flavor of the information that the matched CPS can provide about the measurement error process. We focus on discrete variables, for in this case the measurement error process is completely described by a finite dimensional probability matrix.[11] Our analysis is subject to two limitations. First, some of the patterns observed in the data may be due to false matches. Second, we do not relate the occurrence of measurement errors to the characteristics of the respondent. This information is only available in the CPS starting with 1990.[12]

5.1. *Reporting errors in sex and race*

We begin by presenting the results of two exercises. In the first we match individuals on the person identifier, age and race, and we consider sex changes between two March surveys. In the second we match individuals on the person identifier, age and sex, and we consider race changes.

[11] Measurement errors in a discrete variable X do not satisfy the assumptions of the classical error-in-variables model. In particular, the conditional distribution of the measurement error depends on the true value of X and does not have mean zero (see, e.g., Aigner, 1973). Further, measurement errors are negatively correlated with the true value of X.

[12] Simple tabulations for that year show that 62 percent of the respondents are women. Of these, 46 percent are female household heads, while 48 percent are spouses.The ratio of male to female respondents varies little with age, except for the late ages (after age 70), where women are much more likely to be the respondent.

Table 7
Percentage race changes by race and age group

Year 1 age	Year 1 race	Year 2 race		
		White	Black	Other
0–14	White	99.7	0.1	0.2
	Black	0.8	99.1	0.1
	Other	3.3	0.2	96.5
30–14	White	99.8	0.1	0.1
	Black	0.7	99.3	0.1
	Other	2.6	0.2	97.2

Sex changes are strongly related to the age of a matched person. They are highest (1.6 percent) for those aged 0 to 14 in the first survey, and lowest for mature and older people aged 30 to 65 (0.3 percent).

On the other hand, the distribution of race changes depends on reported race in the first survey but not on age. As shown in Table 7, the frequency of race changes for people aged 30 to 65 initially classified as whites is negligible (0.2 percent). It is a little higher for people initially classified as blacks (0.7 percent). It is substantial, however, for people initially classified as other nonwhites (2.8 percent). Interestingly, for both blacks and other nonwhites, over 90 percent of the race changes are towards white. One way of thinking about measurement errors in this case is in terms of a gravity model, where a person reports the median attribute of the population instead of its true attribute. In contrast with the standard measurement error model, this kind of measurement error is variance-reducing rather than variance-enhancing.

5.2. Reporting errors in age

Although age varies deterministically through time, the imperfect coincidence of the interview dates in the two surveys implies that no age difference and two-year differences are possible although unlikely values.

First consider all individuals that are matched on the person identifier, sex, race, and age, with age changes allowed to vary between zero and two. In Fig. 7 we present, separately by sex and age, the percentage reporting zero-, one-, and two-year age changes. Among those aged 0 to 65 in the first year, no age changes occur for 3.3 percent of the men and 3.4 percent of the women. Interestingly, this percentage tends to rise with age, especially for women. Further, there is clear evidence of heaping at certain frequencies, typically multiples of five.

Now consider people aged 0 to 65 in the first year, who are matched on the person identifier, race, and sex. Age changes greater than two or negative are

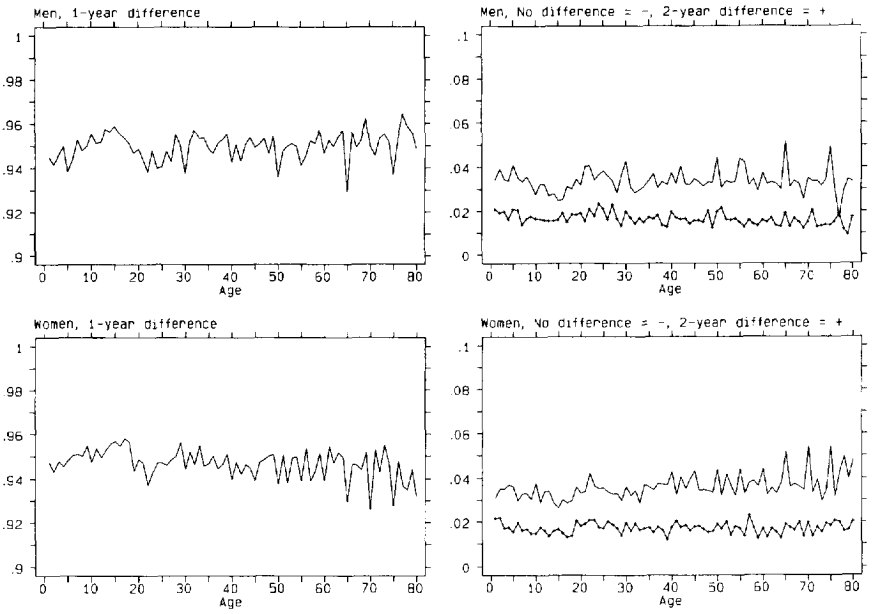

Fig. 7. Age changes by sex and age.

reporting errors. We find that negative age changes occur for 1.3 percent of both men and women, and the average age reduction is 8.4 years for men and 9.5 years for women. Age changes greater than two years occur for 1.2 percent of the men and 1.4 percent of the women, with an average age increase of 10.8 years for men and 11.2 years for women.

5.3. *Reporting errors in schooling*

Changes in years of schooling completed are necessarily nonnegative. Further, changes greater than one year ought to be considered as reporting errors.

In Fig. 8 we present reported differences in schooling by sex, age, and schooling level. Except for the top-left panel, data have again been smoothed somewhat using Cleveland's (1979) loess. The percentage of men and women reporting zero differences rises from 60 to over 90 percent between age 20 and age 25, and reaches 95 percent after age 40. This percentage is lowest for those who did not finish high-school or college, and is highest for high-school graduates, which may be consistent with individuals completing their degree at later ages.

The two panels on the right-hand side of the figure present some evidence on reporting errors. These are highest at younger ages (about 6 percent at age 15), and decline afterwards stabilizing at about 3 percent after age 35. Interestingly,

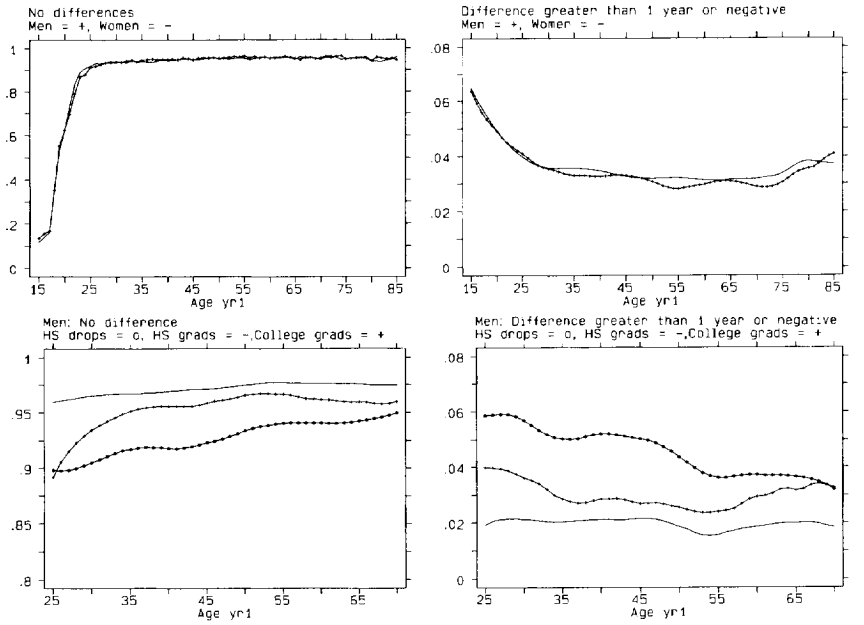

Fig. 8. Differences in reported education.

error rates are highest for those who did not finish high-school or college, and are lowest for high-school graduates.

6. Conclusions

Matched CPS files appear to be an interesting vehicle for analyzing both short-run labor force dynamics and the nature of the measurement error process in survey data. Their main advantages are the size, and the fact that they span several cohorts and maintain a relatively stable distribution by age and other person characteristics.

If the matched CPS is used to analyze short-run dynamics, it is important to understand both the process that determines attrition from the sample and the nature of measurement errors in the variables of interest. Our results show that the matched and the unmatched are different in important dimensions. In particular, attrition from the CPS is concentrated among the younger, and appears to be the result of household and person mobility due to decisions about schooling, family formation, and job search. Although attrition from the survey and new entry do not affect the representativeness of the CPS as a cross-section, the loss to the sample being relatively similar to the new entrants, they

may affect behavioral relationships estimated from the matched data. Somewhat surprisingly, we find that while selection on the matched does bias measures of participation and wages, especially for men, no major bias appears in the estimates of transitions between labor force states after controlling for sex, age, and labor force status at the time of the first survey.

References

Aigner, D.J., 1973, Regression with a binary independent variable subject to errors of observation, Journal of Econometrics 1, 49–60.

Baltagi, B., 1985, Pooling cross-sections with unequal time-series length, Economics Letters 18, 133–136.

Berg, G.J. van den, M. Lindeboom, and G. Ridder, 1991, Attrition in longitudinal panel data, and the empirical analysis of dynamic labor market behavior, Institute of Economic Research Research memorandum no. 427 (University of Groningen, Groningen).

Bjorn, E., 1981, Estimating economic relations from incomplete cross-section/time-series data, Journal of Econometrics 16, 221–236.

Cleveland, W.S., 1979, Robust locally-weighted regression and smoothing scatterplots, Journal of the American Statistical Association 74, 829–836.

Deaton, A., 1985, Panel data from time series of cross-sections, Journal of Econometrics 30, 109–126.

Fuller, W.A. and G.E. Battese, 1974, Estimation of linear models with crossed-error structure, Journal of Econometrics 2, 67–78.

Hausman, J.A. and D.A. Wise, 1979, Attrition bias in experimental and panel data: The Gary income maintenance experiment, Econometrica 47, 455–473.

Murphy, K.M. and F. Welch, 1992, The structure of wages, Quarterly Journal of Economics 57, 285–326.

Peracchi, F. and F. Welch, 1994, Trends in labor force behavior of older men and women, Journal of Labor Economics 12, 210–242.

Pitts, A., 1988, Matching adjacent years of the Current Population Survey, Mimeo. (Unicon Research Corporation, Santa Monica, CA).

Poterba, J.M. and L.H. Summers, 1986, Reporting errors and labor market dynamics, Econometrica 54, 1319–1338.

Wansbeek, T. and A. Kapteyn, 1989, Estimation of the error-component model with incomplete panels, Journal of Econometrics 41, 341–361.

Welch, F., 1993, Matching the Current Population Surveys, Stata Technical Bulletin 12, 7–11.