A Cautionary Note on Using (March) CPS and PSID Data to Study Worker Mobility*

Gueorgui Kambourov[†]
University of Toronto

Iourii Manovskii[‡]
University of Pennsylvania

Abstract

The monthly Current Population Survey (CPS), with its Annual Demographic March supplement, and the Panel Study of Income Dynamics (PSID) are the leading sources of data on worker reallocation across occupations, industries, and firms. Much of the active current research is based on these data. In this paper, we contrast these datasets as sources of data for measuring the dynamics of worker mobility. We find that (i) (March) CPS data is characterized by a substantial amount of noise when it comes to identifying occupational and industry switches; (ii) March CPS data provides a poor measure of annual occupational mobility and, instead, most likely measures mobility over a much shorter period; (iii) (the changes in) the procedure to impute missing data has a dramatic effect on the interpretation of the CPS data in, e.g., the trend in occupational mobility. The most important shortcomings of the PSID are the facts that (i) occupational and industry affiliation data is available in most years at an annual frequency; (ii) the PSID's sample, by design, excludes immigrants arriving to the U.S. after 1968; (iii) the Retrospective Occupation-Industry Files with reliable occupation and industry affiliation data are available only until 1980.

JEL classification: E24, J62.

Keywords: Current Population Survey, Panel Study of Income Dynamics, Occupational Mobility, Worker Mobility, Coding Error, Imputation Procedures.

*First version: March 26, 2004. This version: May 5, 2010. We would like to thank an anonymous referee for detailed comments and suggestions, and Irina Telyukova for excellent research assistance. We gratefully acknowledge support from the Social Sciences and Humanities Research Council of Canada Grants #410-2007-0299 and #410-2008-1517 and the National Science Foundation Grants No. SES-0617876 and SES-0922406.

†Department of Economics, University of Toronto, 150 St. George St., Toronto, ON, M5S 3G7, Canada. E-mail: g.kambourov@utoronto.ca.

‡Department of Economics, University of Pennsylvania, 160 McNeil Building, 3718 Locust Walk, Philadelphia, PA, 19104-6297 USA. E-mail: manovski@econ.upenn.edu.

1 Introduction

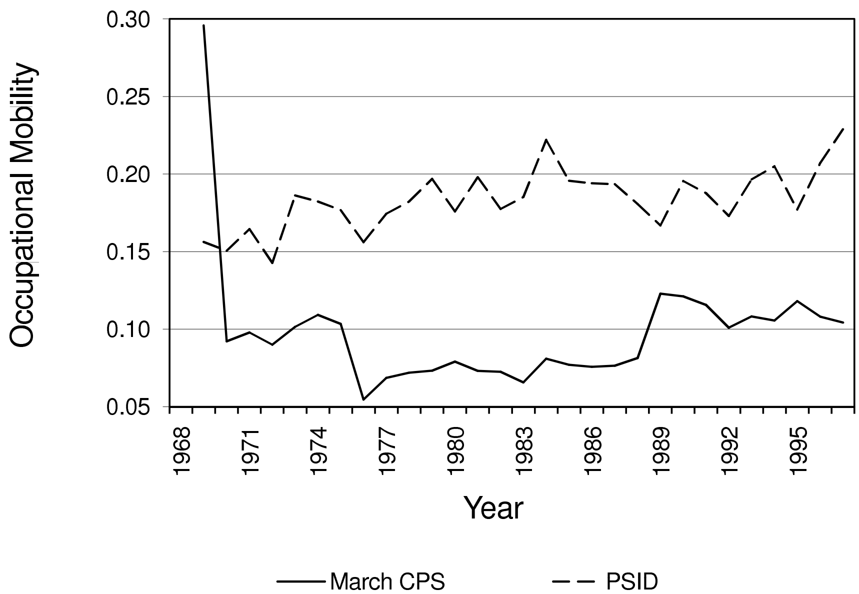
There is considerable research interest in studying worker mobility across occupations, industries, and employers. The existing longitudinal studies of worker mobility in the U.S. are typically based on the Current Population Survey or the Panel Study of Income Dynamics. In this paper, we describe some pros and cons of using each of these two data sets, especially for the study of changes in mobility over time. Each of the two data sets has a distinct set of advantages and disadvantages that affect the comparability of findings between them and make them uniquely useful for understanding particular economic phenomena. While our discussion is focused on these two surveys, the data issues we describe are not unique to these sources of data and apply, for example, also to data from the Census of Population, the National Longitudinal Surveys, or the Survey of Income and Program Participation, and to most survey data on worker mobility collected in other countries.

The Current Population Survey (CPS) is the source of the US official labor market statistics. It is also the primary source of the data used by researchers to study worker flows across occupations, industries, and employers. The CPS appears particularly well suited for this purpose because it represents a sequence of monthly cross-sectional samples of a very large size representative of the US population. Moreover, there is a limited panel dimension because individuals are interviewed in several consecutive months allowing for studying short but high-frequency labor market histories. Furthermore, in March of every year a supplement is administered to all members of the survey in which they are asked about their labor market information in the previous calendar year allowing for the study of worker transitions across labor market states over a longer period. Despite the many appealing features of the Current Population Survey, we identify in this paper several reasons why the statistics computed from its data must be interpreted with caution.

The Panel Study of Income Dynamics (PSID) is a much smaller but a truly longitudinal survey. The PSID started in 1968 with a nationally representative sample (after using weights to account for the oversample of economically disadvantaged households) of about 5,000 U.S. households. Since then, the PSID has attempted to follow all individuals from those households in addition to all their current co-residents (spouses, cohabitors, children and anyone else living with them). When a sample family separates, both the heads and spouses are followed. Moreover, the study's design ensures that children born to these families are a nationally representative sample of newborns; therefore, these children also become part of the sample and are followed when they leave home. The PSID's rules for tracking individuals and family units over time have led to a continuous representation of the U.S. population in terms of all demographic changes other than immigration (Duncan and Hill (1989)). The PSID follows individuals over a long period of time as they locate and change occupations, industries, and employers. Moreover, as we discuss below, the use of the PSID Retrospective Occupation-Industry Files allows us to minimize the amount of coding error in occupational and industry transitions of individual workers.

Consider Figure 1 that depicts occupational mobility at the three-digit level obtained by Kambourov and Manovskii (2008) from the Panel Study of Income Dynamics (PSID) and March CPS. Occupational mobility in each data set is computed on identical samples of male workers aged 23-61, who are not self- or dual-employed, and are not working for the government.

Figure 1: Occupational Mobility in the US, PSID and March CPS.



The difference in identified levels of mobility in these data sets is large. The annual three-digit mobility level is around 10% in the March CPS and 18% in the PSID. The occupational mobility measured on the March CPS declines even further to about 7% once individuals with imputed observations are dropped from the sample after 1988 (a common practice, the effects of which we discuss below). In what follows, we will suggest that one reason for this difference is that the March CPS data does not identify annual mobility rates. Instead, it likely measures mobility over a two- to three-month period. In addition, we will argue that CPS data on occupation and industry affiliations is characterized by a surprisingly large measurement error. We will also argue that (changes in) the imputation procedure has a strong effect on the statistics obtained from the CPS data and on its interpretation.

While these arguments suggest that the PSID data might be preferred for longitudinal studies of mobility, it does have its share of problems. Most important among those are the facts that occupational and industry affiliation data is available in most years only at an annual frequency, which makes it impossible to identify higher frequency transitions; measures of employer mobility are noisy; the occupational classification has not been changed from 1968 to 2001 which, on the one hand, makes many statistics comparable over time but, on the other hand, leads to potentially important biases because the classification becomes more out-dated over time. Moreover, the PSID's sample, by design, excludes immigrants arriving to the U.S. after 1968. While some attempts were made to address this issue, in most years the findings on the PSID can be generalized only to the non-immigrant population. Finally, the Retrospective Occupation-Industry Files with reliable occupation and industry affiliation data are available only until 1980, which makes the analysis of mobility, especially at the individual level, more difficult and less reliable in subsequent years.

The paper is organized as follows. In Section 2 we discuss the sources and the magnitude of coding error in occupational affiliation data in the PSID and CPS. In Section 3 we argue that the March CPS measures mobility at a much higher frequency than the annual. In Section 4, we discuss the effects of the data processing and imputation procedures on the interpretation of occupational mobility data in the CPS. In Section 5, we discuss some of the shortcomings of the PSID data relevant for the study of worker mobility. Section 6 concludes the paper.

2 Spurious Transitions Induced by Occupation and Industry Coding

2.1 The Origins of Coding Error

The biggest amount of measurement error in occupation and industry affiliation is generated at the coding stage, which the following experiment summarized in Mathiowetz (1992) illustrates. Reports of occupations obtained in interviews of employees of a large company were checked against the descriptions contained in company records. This was done in two ways. First, the coders were asked to compare simultaneously the two descriptions and to code them as being in agreement if the two sources could result in the same three-digit classification. The procedure resulted in a disagreement rate of 12.7%. Second, the coders independently coded the two descriptions at the one- and three-digit levels. The comparison of the independently assigned codes resulted in a disagreement rate of 48.2% at the three-digit level and 24.3% at the one-digit level. These results, summarized in Table 1, indicate that by far the largest amount of error in occupational or industry affiliation data is generated at the coding stage. This remains true even if one uses a much more aggregated one-digit occupational classification. In addition, Table 1 indicates that the magnitude of the error rises when workers are asked to describe their occupations a year before the interview. This will be important when we discuss the March CPS where workers are asked about their occupations at the time of the interview and in the previous year. These findings are consistent with those reported by Mellow and Sider (1983).

Table 1: The Origins of Coding Error.

	Proportion Disagreement between Interview and Company Records	
Method of Comparison	Current Occupation	1 Year Prior to Interview
Joint coding of interview and company record:	.127 (.017)	.184 (.020)
Independent coding of interview and company record:		
3-digit code	.482 (.026)	.510 (.026)
1-digit code	.243 (.022)	.259 (.022)

Note: Standard errors are in parentheses.
Source: Mathiowetz (1992).

2.2 Coding Error in the PSID

To further highlight the effect of the coding procedure on measurement error in occupational affiliation, consider two coding techniques used by the PSID. The PSID has used the 1970

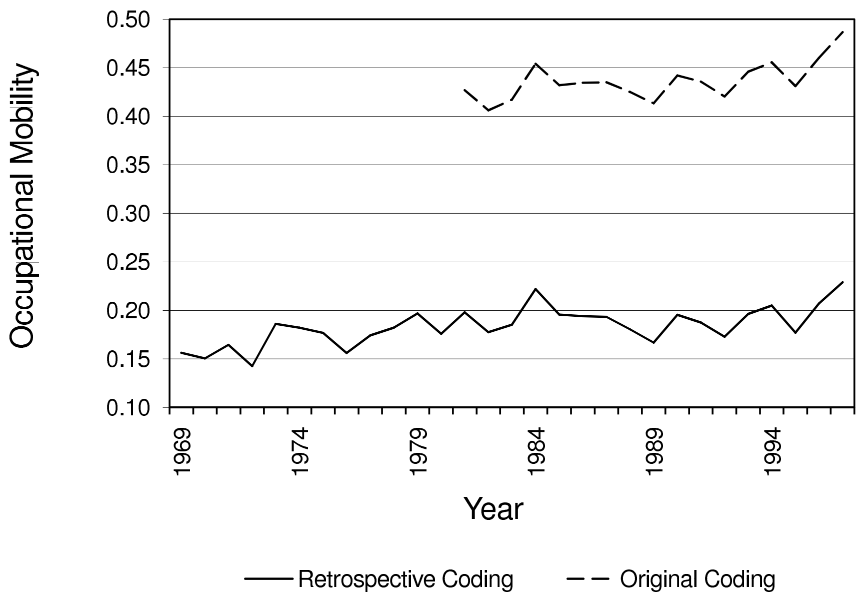
Census occupation and industry classification from 1968 on. However, one-digit occupation codes were used in 1968-1975, two-digit occupation codes in 1976-1980, and three-digit occupation codes after 1981. The industry affiliation was coded at a two-digit level in 1971-1980, and at a three-digit level after 1981.

In 1996 the PSID started working on the 1968-1980 Retrospective Occupation-Industry Files. This work originated as part of the Working Lives and Mortality in an Aging National Cohort project. That project required three-digit occupation and industry codes throughout the course of the PSID. As mentioned above, the PSID did not originally code occupations and industries at the three-digit level prior to 1981. In order to produce the three-digit recode, the PSID pulled out paper materials from its archives containing the written records of the respondents' descriptions of their occupations and industries. These were the same records from which the one and two-digit occupations and industries were coded prior to 1981. Using these records, the PSID assigned three-digit 1970 Census codes to the reported occupations and industries of household heads and wives for the period 1968-1980. The work was completed in 1999, when the PSID released the Retrospective Occupation-Industry Supplemental Data Files. Using the Retrospective Files, we create a series of consistent three-digit occupational codes that runs from 1968 till 1997.

Surprisingly, at first we find a significant degree of disagreement between the originally assigned PSID occupation and industry codes and the codes assigned to the same individuals in the Retrospective Files. Consider, for example, the two-digit occupational mobility for the 1976-1980 period. During this period the PSID provides the originally assigned occupation (industry) codes as well as the codes reassigned in the Retrospective Files. One would expect the levels of occupational mobility computed on these two series to be similar, if not exactly the same, since both are based on the same raw information − the respondent's description of his or her occupation contained in the PSID interview records. Any difference must come from the way the original information contained in those records was transferred into an occupation code. One finds, however, that the level of occupational mobility in the Retrospective Files during the 1976-1980 period is roughly half, at approximately 14%, of the rate of mobility obtained on the originally coded occupations that is approximately 27%. Similarly, Figure 2 depicts occupational mobility of over 40% at the three-digit level obtained from the originally coded PSID data (dotted line) and less than half at around 18% from the Retrospective Files (solid line).[1]

[1]The Retrospective Files are available from 1968 till 1980. After 1980, the series is adjusted using an estimate of the coding error. See Kambourov and Manovskii (2008) for a detailed description of the procedure.

Figure 2: Occupational Mobility in the US, 1969-1997, PSID, Three Digit Level, Original and Retrospective Coding.



Why is occupational mobility so much lower when computed using the Retrospective Files? The difference between the originally and the retrospectively assigned occupation and industry codes was caused by differences in the methodology employed by the PSID in constructing these data. When originally coding the occupation (industry) data, the PSID coder could not compare the current year description to the one in the previous year. As a result, for a respondent who is in the same occupation (industry) in both years, similar occupational (industry) descriptions could end up being coded differently. This was not the case with the constructed Retrospective Files where, as reported in the PSID (1999), "To save time and increase reliability, the coder coded all occupations and industries for each person across all required years before moving on to the next case." Thus, in constructing the Retrospective Files, the coders had access not only to the respondents' description of their current occupation (industry), but also to the description of their past and future occupations (industries). This allowed them to compare these descriptions, decide whether they are similar, and assign the same occupational (industry) code where appropriate. The use of this procedure suggests that the PSID Retrospective Files provide reliable data for identifying occupational and industry switches, which is confirmed formally in Kambourov and Manovskii (2009b).

2.3 Coding Error in the (March) CPS

Before discussing measurement error issues in the (March) CPS data, a brief description of the data is in order. The Current Population Survey (CPS) is administered by the Bureau of the Census under the auspices of the Bureau of Labor Statistics (BLS). Currently, about 65,000 households, a nationally representative sample, are interviewed monthly. Each household is interviewed once a month for four consecutive months one year, and again for the corresponding four months a year later, resulting in 8 total months in the survey. Each month, a new cohort (rotation group) is added to the survey, and the one that completed its eighth interview is permanently retired. Thus, eight rotation groups are interviewed in any given month.

Table 2: CPS Structure: 4 Different March Cohorts.

Cohort	Dec	Jan	Feb	Mar	Apr	May	Jun	–	Dec	Jan	Feb	Mar	Apr	May	Jun
I	1	2	3	4				–	5	6	7	8			
II		1	2	3	4			–		5	6	7	8		
III			1	2	3	4		–			5	6	7	8	
IV				1	2	3	4	–				5	6	7	8

Each month in the sample, adults (currently defined as 15 years of age and older) are asked about their labor force activity for the week prior to the survey, as well as about their employment status, occupation, and industry. Occupational and industry codes are then assigned to the descriptions of the current occupation and industry according to a *dependent coding* technique detailed in the following subsection. However, it is important to point out that the monthly CPS uses dependent coding only after 1994.

In March of each year, all the cohorts of workers present in the basic CPS sample (see Table 2 for a schematic description) are administered a supplemental questionnaire in which they are asked to describe their main (longest) job in the previous year. Occupational and industry codes are then assigned to the descriptions of the job the previous year according to a similar dependent coding technique as in the monthly CPS after 1994. The dependent coding has been used in the March CPS since 1970.

2.3.1 CPS Dependent Interviewing and Coding Procedure

In the first and the fifth month in the CPS sample, respondents are asked to describe their usual activities and duties. Based on these descriptions, coders assign occupational codes to these descriptions. In each of the months 2-4 and 6-8 the individual is asked the following questions:

1. Last month, it was reported that you worked for (company name). Do you still work for (company name)?
   Yes → Ask next question.
   No [4.8%] → Skip to *independent* occupation questions.
2. Have the usual activities and duties of your job changed since last month?
   No → Ask next question.
   Yes [1.9%] → Skip to *independent* occupation questions.
3. Last month you were reported as (previous month's kind of work performed) and your usual activities were (previous month's duties). Is this an accurate description of your current job?
   Yes → Use dependent coding.
   No [2.3%] → Ask *independent* occupation questions.

The percentages in square brackets provide the percent of respondents giving a corresponding answer during the test of the 1994 CPS re-design that introduced dependent coding into the monthly CPS reported in Polivka and Rothgeb (1993).

Even though some coding error is eliminated by using this sequence of questions, a significant amount of noise in individuals' occupational (and industry) affiliation still remains. First, everyone who changes his or her employer is coded independently (see question 1). The fraction of individuals switching employers is substantial and certainly not all of them switch their occupation (less than a half according to the evidence from the Retrospective PSID files). Second, those who remain with the same employer and claim a change in their duties are also independently coded (see question 2). While a change in duties might be due to an occupational switch, many changes in duties are clearly consistent with a worker remaining in the same occupation.[2] Finally, in order to get to question 3 the respondent

[2]At issue here is the level of aggregation the researcher is interested in. An economist in an academic department might serve in various years on a hiring committee, as a graduate or an undergraduate chair, or as a department head. All of these transitions involve a change of duties. For the analysis of some questions, these changes of duties should be taken as a primitive. For others, it is more important that the person remains an economist, and that all these changes in duties are consistent with being an university economist. If the researcher is interested in mobility at an even more aggregated level, it would be appropriate to classify this worker as remaining in the same broad occupational category even if he had become a librarian. On the

must have answered that neither her employer nor her duties at her job have changed since the previous month. Therefore, it is hard to imagine a case where someone would then claim that the type of job and activities reported for the previous month's job is not an accurate description of their current job. Nevertheless, a significant fraction of respondents fall into this category and are independently coded. The likely reason for this is that over half the CPS data is collected through proxy interviews (persons responding for other household members) and the respondent, who can be any household member older than 15, often varies from one month to the next (Rothgeb and Cohany (1992)).

Consider the following back-of-the-envelope calculation which illustrates the magnitude of the coding error made despite the dependent coding technique. Suppose the true level of mobility is 2% a month. Given that for 9% of workers occupations are coded independently from month to month, occupational descriptions for 7% of workers who did not switch occupations would end up being coded independently. With a coding error of about 50%, consistent with the findings in Mathiowetz (1992), 3.5% of workers will be identified as switching occupations when no such switch has actually occured, i.e., the measured mobility would be around 5.5% a month instead of the true 2%. Indeed, we will show below that monthly measures of occupational mobility from the March CPS are in excess of 5%.[3]

The discussion above implies that the monthly CPS is characterized by a substantial amount of noise when it is used to identify occupational switches, even after the introduction of dependent coding after 1994. The March CPS uses a similar procedure since 1970, although the questions compare descriptions of the current job (obtained in the Basic CPS interview) and the longest job held the previous year. The introduction of dependent coding into March CPS decreased the amount of coding error as is evident from a substantial decline in mobility between 1969 and 1970 observed in Figure 1. The discussion above implies, however, that a substantial amount of coding error remains.

Recognizing the magnitude of the coding error is important for drawing the correct conclusions from the data. The following examples illustrate.

other hand, even the fairly disaggregated three-digit occupations may not be sufficiently fine to capture the relevant changes in duties. For example, a family physician who becomes a specialist in internal medicine at a hospital may substantially change his duties to warrant a new occupational title, but even the three-digit classification does not permit such a change.

[3]Unfortunately, we have no way of knowing what fraction of independently coded workers are true occupational switchers from one month to the next. Let's suppose that 50% of job switchers are occupational switchers as are 50% of job stayers who claim that their duties have changed. It would seem likely that very few workers in the same job with the same duties but who disagree with their last period's description of occupations are actually occupational switchers. In this example, true mobility is $0.5 \times (4.8 + 1.9) = 3.35\%$ while an additional $0.5 \times (9 - 3.35) = 2.825\%$ would be erroneously classified as switchers, assuming a 50% rate of coding error. Even with the rate of coding error as low as 30%, over one third of identified occupational switches would be spurious.

1. In the first and fifth months, occupations are coded independently. Thus, there is a substantial error in identifying the genuine occupational affiliation of the workers during those months. Dependent coding helps, to some extent, lower the extent of spurious transitions identified during the three subsequent interviews, but it does not help identify true occupational affiliations. This is important for studies that require knowledge of the number of workers in a given occupation, e.g., Fallick (1993), Lee (2005), and Lee and Wolpin (2006). These studies also attempt to evaluate to what extent sectoral reallocation is driven by differences in average wages across sectors. The presence of measurement error in the occupational classification is likely to bias downward the observed differences in average wages.

   This bias is also important for the literature measuring the inter-industry wage differentials. For example, Krueger and Summers (1988) make an effort to correct for the measurement error in worker transitions across industries but ignore the fact that the one-digit occupational dummies they use to control for workers' skills are measured with error and might not capture true occupational effects. Similarly, Dickens and Katz (1987) document high correlations of the average log wages of one-digit occupations across industries. This may not be informative if the occupational dummies they use do not capture true occupational affiliations because of the high rate of coding error.

2. Coding error inflates the measures of gross worker mobility. At the same time, it introduces a downward bias into the measured net mobility which is typically defined as one half of the sum of the changes in occupational employment shares between two points in time. To see this, consider the following stylized example. Suppose there are only two occupations, A and B, with a large population of workers equally split between them. Now suppose that the coding error is 50% so that a worker in, say, occupation A is equally likely to be classified as working in either occupation. In this world, measured gross mobility is equal to 50% regardless of the true underlying mobility. At the same time, measured net mobility is always zero regardless of the extent of the true mobility. This implies that the measured distance between gross and net mobility is affected by the extent of the coding error. Yet, the existing studies that rely on the measures of these differences to identify the driving forces for worker mobility (e.g., Artuc, Chaudhuri, and McLaren (2007), Murphy and Topel (1987), Jovanovic and Moffitt (1990), and Moscarini and Vella (2002)) make no attempt to

correct for this bias.[4]

3. Moscarini and Thomsson (2007) use the occupational mobility data from the monthly CPS to estimate the magnitude of worker transitions across jobs. They note that around 5% of the observations in the monthly CPS have missing answers to the question of whether a worker switched employers from one month to the next. However, for many of these workers the occupational codes are available. They assume that all the switches identified after 1994 using the data with non-missing answers to the sequence of the dependent coding questions are genuine (even the ones that are effectively independently coded). Next, they propose several filters that identify as erroneous some of the occupational switches among workers whose answers to the sequence of dependent coding questions are missing.[5] Finally, they compute the fraction of occupational switches with valid data that are accompanied by an employer switch and apply the same ratio to observations with occupational switch but missing employer switch data to impute the overall level of job-to-job mobility. This procedure crucially relies on assuming that all occupational switches identified by the CPS's dependent coding technique are genuine. If this is not the case, the adjustment may bias the estimated rate of job-to-job mobility. Obtaining reliable measures of job-to-job flows is important because they are of central interest to the calibration and estimation of equilibrium search models of the labor market.

In conclusion of this Section, we would like to emphasize that the introduction of the dependent coding procedures in the March CPS in 1970 and in the monthly CPS in 1994 has substantially reduced the number of spurious occupational and industry switches. Nevertheless, some noise remains and should be taken into account in interpreting the findings based on these data. As a practical suggestion for the design of the dependent coding procedure, we think that it might be better implemented as follows. The respondents should be asked only one question regardless of whether they switched their employer or not: "Last month you were reported as (previous month's kind of work performed) and your usual activities were (previous month's duties). Is this an accurate description of your current

[4] The distance between the gross and net mobility is potentially informative because the models of worker mobility based on fluctuating demand for various sectors of employment (e.g., Lucas and Prescott (1974)) imply net mobility that is high relative to gross mobility, while the models based on matching considerations between workers and firms (e.g., Jovanovic (1979)) or workers and occupations (e.g., McCall (1990)) imply high gross mobility relative to net mobility.

[5] Liu and Trefler (2008) apply a similar filter to the annual data from matched March CPS surveys to minimize the effects of the coding error in their study of the effects of changes in international trade on worker mobility across occupations and industries.

job?” If the answer is “Yes,” last month’s occupational code should be carried forward. If the answer is “No,” a new description should be taken from the respondent and both the last month’s and the current description should be made available to the coder. Looking at both descriptions, the coder should decide whether the reported change in duties warrants a change in the occupational code (depending on the occupational classification used).

3 Does the March CPS Measure Annual Mobility?

A number of important contributions in the literature (e.g., Moscarini and Vella (2003), Lee (2005), and Lee and Wolpin (2006)) have interpreted measures of mobility from the March CPS as the rate of mobility between two consecutive years. We argue in this section that this interpretation may not be correct. The knowledge of the time horizon over which March CPS measures mobility is crucial for the estimation and calibration of structural macro/labor models to be consistent with their model period and choices of the other parameter values.

Instead of relying on the March CPS data to measure annual mobility, one can compute it directly by matching the individuals present in the CPS in March of two consecutive years. Appendix Table A-1 reports the matching rates in the monthly 1976-2004 CPS for various time periods − from one month to one year. We provide two different matching rates.[6] The first is the fraction of individuals who can be matched based on the formal identifiers provided in the dataset. The second matching rate, reported as a fraction of those already matched by the first method, excludes all those who have been matched but show a different gender, race, or age more than three years apart. Table 3 reports the corresponding occupational (and industry) mobility on a sample of matched individuals as reported in columns (2) in the Appendix Table A-1. We do not report the mobility rates when the matching rates are very low or for years when the occupational classification changed (1983, 1992, and 2003).

We find that this truly annual measure of occupational mobility is in excess of 40% per year. In particular, over the 1994-2004 period, the March-to-March annual occupational mobility is 47%. The fact that the level of mobility obtained using the March CPS is six times smaller than the genuinely annual mobility measure in the matched Basic CPS indicates that the March CPS does not identify annual occupational mobility. The difference is not due to the fact that occupational mobilty is dependently coded in the March CPS but

[6] See Madrian and Lefgren (2000) and Feng (2001) for a detailed discussion about matching individuals in the monthly CPS.

Table 3: Occupational and Industry Mobility in the US, 1976-2004, Monthly CPS, Three Digit Level.

Year	December-March		January-March		February-March		March-March	
	OCC	IND	OCC	IND	OCC	IND	OCC	IND
1976								
1977	35.8	22.4	33.7	20.5	31.8	18.7	43.3	28.9
1978			31.6	18.4	29.0	16.2		
1979	34.1	20.0	33.1	18.6	30.5	16.8	41.7	27.2
1980	34.5	20.1	33.0	18.6	30.8	17.2	42.0	26.5
1981	34.2	20.9	32.6	19.6	30.3	17.9	41.8	27.1
1982	35.2	21.4	31.8	18.7	30.4	17.7	41.8	26.7
1983			34.1	19.1	31.8	17.6		
1984	37.1	22.0	35.1	20.6	33.0	18.9	44.0	27.6
1985	39.0	23.1	36.6	21.7	34.3	19.8	44.2	28.9
1986	38.9	23.9	36.7	22.1	34.8	20.5		
1987	37.5	22.0	35.7	20.5	33.1	18.2	45.9	29.9
1988	39.2	22.8	35.1	20.2	33.1	19.0	44.8	28.4
1989	38.1	22.4	35.9	20.5	33.6	18.7	45.0	28.7
1990	39.7	23.6	37.3	22.1	34.8	20.2	45.7	29.4
1991	38.3	22.7	37.0	21.6	34.2	19.4	44.7	28.9
1992			35.6	20.4	33.6	18.8		
1993	38.9	23.2	37.0	21.5	34.6	19.8	44.7	28.4
1994	39.8	24.5	11.0	7.5	5.7	3.8	46.7	29.6
1995	13.6	9.7	11.5	8.6	6.6	5.2	46.2	31.1
1996	15.0	11.2	10.3	7.5	5.9	4.4		
1997	15.0	11.4	10.1	7.6	5.6	4.3	46.1	31.4
1998	12.7	9.3	9.1	6.8	4.9	3.6	46.6	31.4
1999	13.1	9.6	9.1	7.1	5.2	3.9	46.9	32.1
2000	12.9	9.9	9.4	7.5	5.4	4.2	46.8	32.9
2001	13.5	9.7	9.6	7.2	5.1	3.7	47.1	32.9
2002	12.7	9.5	9.0	6.9	5.0	3.7	47.5	32.4
2003			8.5	6.4	4.8	3.6		
2004	12.7	9.5	8.5	6.3	4.8	3.6	47.9	34.1

Notes: The table reports the occupational (OCC) and industry (IND) mobility from the monthly CPS for the periods December-March, January-March, February-March, and March-March. The mobility rates are computed on a sample of matched individuals as reported in columns (2) in Table A-1. We do not report the mobility rates when the matching rates are very low or for years when the occupational classification has changed (1983, 1992, and 2003).
Source: Authors' calculations from the monthly CPS.

is independently coded when we use the monthly CPS to calcucate the March-to-March annual occupational mobility. A comparison with the PSID illustrates this. The level of mobility in the matched March-to-March CPS is virtually identical to the one obtained from the noisy originally coded PSID data. The most reliable estimates of the rate of occupational mobility from the PSID Retrospective Files are nevertheless twice as high as the estimates based on the March CPS (as can be seen in Figure 1).

Of course, there is no reason for the March CPS to measure mobility between two points in time that are one year apart. Indeed, the occupational code provided in the March CPS refers to the longest job in the previous calendar year. Thus, one might expect that on average it refers to the occupation held in the middle of the previous year so that by comparing the occupation in the March CPS to the one in the Basic March CPS a mobility over only a nine month period is identified. Even this does not seem to be the case, however. Instead, we suggest that the March CPS measures mobility between March of this year and the very end of the previous year.

We put this hypothesis to the test by computing occupational mobility by matching individuals who are in the CPS sample and report occupations in January, February, and March of the same year, as well as in December of the previous year. We now focus our analysis on the 1994-2004 period, when the CPS implemented its most reliable dependent interviewing technique. The corresponding December-March, January-March, and February-March occupational mobility is 13.5%, 9.6%, and 5.4%, respectively. The mobility measures from the March CPS are, in magnitude, close to the two- and three-month mobility found in the monthly CPS. Therefore, we conclude that the March CPS data suffers from a severe time aggregation problem in worker responses.

4 The Impact of CPS Imputation and Data Processing Procedures

Perhaps, the most important issue in using (March) CPS data in longitudinal studies of worker mobility across occupations and industries is the imputation of missing records and the changes in the data processing procedures by the CPS.[7] In this section we focus the

[7] Other potentially important measurement issues not discussed in this paper are: (1) The occupational and industry classifications changed in 1968, 1971, 1983, 1992, and 2003. Upon each of those changes some new occupations are introduced, while some other previously separate occupations are aggregated. This naturally affects the measured levels of mobility over time; (2) Occupational mobility can be defined only for people employed both in the year of the interview and the previous year. The sample proportion of such people changes substantially over time; (3) Interviewing techniques have changed over time (presumably reducing spurious transitions); (4) The effects of a substantial decline in the response rates after the 1994

analysis on the effect of a change in the method for imputing missing values by the CPS in 1976 and 1989. There are three reasons why a record may be imputed.

1. A respondent was not interviewed by the basic CPS survey in March (either because CPS could not contact the respondent or the respondent refused to answer questions). The overall non-interview rate was quite stable at 4 to 5 percent from the early 1960s until 1994 (U.S. Bureau of the Census (2002)). It increased substantially starting in 1994 to over 8 percent.

2. The nonresponse rate for the March Supplement represents an additional 8 to 12 percent.

3. There might be missing answers to individual questions in the March Supplement or in the basic CPS survey in March. An individual item nonresponse occurs when the respondent does not know the answer or refuses to provide an answer to one or more questions in the survey. For example, in the basic CPS, Heckman and LaFontain (2006) show that the percentage of those who chose not to report earnings was relatively stable at around 15% prior to 1994. After 1994, the earnings nonresponse rose from a low of 24% in 1995 to nearly 34% in 2003.

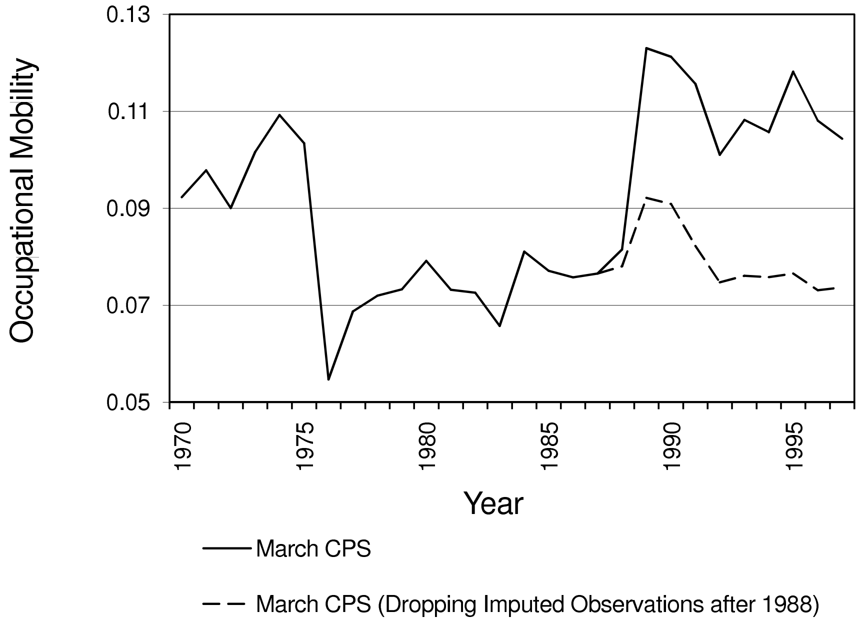
The effects of changes in the imputation method used by the CPS are clearly visible in Figure 3. The solid line represents measured three-digit occupational mobility on the unadjusted March CPS data. It exhibits a sharp drop in 1976 and a sharp increase in 1989. These changes coincide with the change in the imputation technique employed by the CPS.

The CPS uses several techniques for allocating missing observations due to item nonresponse. (1) Relational imputation infers the missing value from other characteristics on the person's record. For example, missing occupation codes are sometimes assigned by viewing the industry codes and vice versa. (2) Longitudinal imputation is used to assign the labor force status by looking at last month's data to determine whether there was a non-allocated entry for that item. If so, the previous month's entry is assigned. (3) "Hot deck" allocation assigns a missing value by picking at random the corresponding observation from records belonging to individuals with similar characteristics. If the entire survey is missing, all the responses are allocated from one donor individual picked by the hot deck procedure.[8]

CPS redesign, discussed below, on measured mobility rates over that period are unclear and warrant an investigation.

[8]See Oh and Scheuren (1980) for a historical background on the CPS hot deck procedure and the U.S. Bureau of the Census (2002) for some details of its current implementation. Hirsch and Schumacher (2004) and Bollinger and Hirsch (2006) describe income imputation hot decks used in the basic CPS.

Figure 3: Occupational Mobility in the US, 1970-1997, March CPS.



The imputation proceeds in a fixed order where first the missing demographic variables are imputed, then the missing labor force status variables, followed by occupations and industries, and finally earnings. Thus, if an occupation is missing, first there is an attempt made to assign it based on the industry data if it is available. If this fails, the occupation is assigned using the hot deck procedure to be the same as for a randomly chosen person of similar age, sex, race, educational attainment, and employment status. In the next step, missing income data is imputed using a hot deck defined by age, sex, race, educational attainment, usual hours worked, and the major occupational group.[9]

[9] The March CPS and basic CPS use somewhat different versions of the hot deck procedure. In the March CPS, the hot deck allocation procedure first tries to match on a fairly rich set of characteristics, at least after 1976. If the match is not found, it proceeds to look for a match on a more and more coarse set of characteristics (until, in the end, a match is guaranteed on sex, age, and race as the only conditioning variables). Thus, for workers with relatively unique characteristics, the donor is likely to be someone with more common characteristics. This observation leads Lillard, Smith, and Welch (1986) to conclude that, "The hot deck trial and error matching process ... yields predictions that are disproportionately near the central tendency of the income distribution; that is, they regress toward modal values and understate dispersion... Whatever one's view of the census understatement among all men, for such professions as doctors and lawyers CPS income by occupation data may be worthless."

In the basic CPS, the match is guaranteed in every category. This is achieved by stocking each cell with a donor who is the most recently encountered by the CPS individual with the characteristics from that cell who had non-missing observations. However, the set of characteristics is fairly coarse and does not include many variables. For example, Hirsch and Schumacher (2004) show that because this set does not contain industry, earnings of workers with missing observations would be replaced by earnings of a donor who likely works in a different industry. This makes the data on inter-industry wage differentials among workers with

Prior to 1989, if the answer to even one of the CPS questions was missing and triggered an imputation, the valid answers to other questions were often removed and replaced with imputed values. For example, suppose an economics professor failed to report his income. The hot deck procedure found a librarian (occupations that belong to the same major group) with similar demographic characteristics and similar hours worked. In this case, not only the value for the missing income is imputed, but also the valid entry for the occupation (economics professor) is erased and is replaced by the occupation of the donor respondent (librarian). This procedure generates spurious occupational transitions from one month to the next and between the basic and the March CPS which are imputed separately. Starting in 1989, this procedure was abandoned and only the missing values are imputed.[10]

While one might expect this to lead to a reduction in measured mobility rates in the March CPS, we instead observe a sharp rise in mobility. The reason is that starting in 1989, not only the new imputation technique was used, but also a completely new processing system was introduced for the March CPS. According to the CPS 2004 Public Use File Technical Documentation,[11] "All of the processing programs were rewritten in 1989, so that not only are the files from 1989 forward based on a somewhat different imputation system, but also reflect a rewritten weighting system, data acceptance program, family relationship edits, and new procedures to match income supplement records to the monthly CPS file. As a result, it is difficult to ascertain whether differences (especially those based on relatively small bases) are the result of imputation or other processing differences between the original and revised files."

The effects of the 1976 change in the imputation procedure are more clear cut. First, starting in 1976, an assignment of missing longest-job information (March CPS supplement) was made from the current job held (basic CPS) if the current job information was available. Prior to 1976, longest-job information was imputed by hot deck using only age

imputed income uninterpretable. In smaller cells, this procedure also has an implication that (1) the same donor may be used to impute many respondents with missing data, and (2) the data for the donor may be dated, often substantially. Bollinger and Hirsch (2006) document that the average donor was sampled almost 9 months prior, and for almost 3% of imputations the observations come from a donor sampled over 5 years prior. Because missing earnings are imputed with donor's values without an adjustment for, e.g., inflation, this results in a substantial understatement of earnings growth.

[10] Lillard, Smith, and Welch (1986) report frequencies in which some information was overridden during imputations for the 1980 CPS: "Although the purpose of allocations is to fill in missing values, it is clear that substitution for observed values occurs as well. For example, 10 percent of respondents' answers to work experience questions were eradicated, a problem that was more serious when finding a match proved difficult. One-third of all blacks had valid work experience answers changed to matched values. But the census eraser will also be busy in thin segments of the white sample."

[11] Available at *http://www.census.gov/apsd/techdoc/cps/cpsmar04.pdf*.

and sex to define "similar person" (U.S. Bureau of the Census (1976)). Second, there was a substantial expansion of categories in the hot deck matching algorithm. These naturally led to a reduction in spuriously imputed occupational and industry transitions and also to a reduction in the actual transitions when the longest job the previous year was imputed with the information from the current job.

It is not clear to us if there is a way to identify which part of Figure 3 is least biased by the imputation and processing procedure in depicting the level of mobility. Even when we ignore the issue of the robust level of mobility, it is not clear how to derive an estimate of the changes in mobility over time that would not be strongly affected by changes in the imputation and processing procedures.

It is also not clear what the preferred method is for dealing with imputed observations. On the one hand, one may want to keep them in to avoid computing aggregate statistics based on a sample with a large proportion of missing data and in an attempt to correct for a possible nonresponse bias caused by non-randomly missing data. On the other hand, one may want to exclude them from the analysis to avoid measuring spurious transitions. Unfortunately, this is not possible, especially before 1988. After 1988 the CPS provides an indicator if the March supplement was imputed, and on Figure 3 we graph occupational mobility on such a sample that excludes all imputed observations after 1988. Many authors, e.g., Stewart (2002), Moscarini and Vella (2003), and Artuc, Chaudhuri, and McLaren (2007) choose to exclude identified individuals with imputed records from the sample after 1988. These individuals are left in the sample before 1988, however, which would seem to make the data before and after 1989 not comparable.

The discussion above has several important implications.

1. Much of the existing literature that argues that there was a decline in the sectoral reallocation of labor over the 1970s is based on the CPS data, and its results seem to be driven by the failure to control for the effects of changes in the imputation methodology. For example, Murphy and Topel (1987) argue that there was a sharp decline in industry mobility in the US between the early 1970s and 1985. Their results are mostly due to a sharp drop in mobility in 1976 coinciding with the change in the imputation procedure. Similarly, Moscarini and Vella (2002) found a substantial decline in occupational mobility in the March CPS data since 1971. Their results are affected by the change in the imputation in 1976, and by dropping imputed observations after 1988.[12]

[12] Although the change in the imputation procedure in the CPS is responsible for the sharp drop in

2. The standard procedure to identify genuine occupation and industry switches from the available noisy codes is to consider an identified occupation (industry) switch to be genuine only if it coincides with some other significant labor market change, such as an employer or an industry (occupation) switch. Kambourov and Manovskii (2009b) have evaluated the quantitative performance of many such filters in the annual PSID data. Moscarini and Thomsson (2007) apply such filters in the basic CPS data. Their performance is affected, however, by the fact that the CPS often uses relational imputation where missing occupation codes are sometimes assigned by viewing the industry codes and vice versa. Thus, such filters are suspect when either occupational or industry codes are imputed.

3. The elimination from the sample of individuals with imputed records after 1988 also strongly affects the cyclical behavior of occupational mobility. Consider the dotted line in Figure 3. This line represents the measured occupational mobility in the March CPS after individuals with imputed records were eliminated after 1988. The graph appears to suggest that worker reallocation became considerably less volatile in the 1990s as opposed to the 1980s. This is, however, an artifact of the elimination of individuals with imputed records from the sample. If these individuals are left in the sample throughout (the solid line), one will find little evidence of a change in the volatility of occupational mobility.

4. As the discussion in Footnote 9 implies, the hot deck procedures used by the basic and March CPS tend to reduce the variance of wages between occupations and industries. This generates a co-movement in average wages across occupations and industries making the identification of, e.g., sectoral shocks difficult. This also affects the measurement of, e.g., inter-industry wage differentials. This effect is on top of the reduction in the variance of sectoral wages induced by the coding error discussed above.

---

occupational and industry mobility in 1976 in the CPS, using data from the Longitudinal Research Database of the Census Bureau and the BLS Labor Turnover Survey, Faberman (2008) finds a secular decline in the job creation and job destruction rates in US manufacturing over the 1947-2006 period. Davis (2008) summarizes the available evidence on a secular decline in the unemployment inflows since the 1970s. This is in contrast to a secular increase in worker mobility across occupations and industries documented in Kambourov and Manovskii (2008).

5 Shortcomings of the PSID for Measuring Worker Mobility

As we have discussed above, the PSID Retrospective Files minimize the coding error in identifying occupational and industry mobility. This enables one to identify the trends in occupational and industry mobility over time. Moreover, the panel dimension of the PSID allows the construction of worker employment histories including measures of occupational, industry, and firm tenure. However, the PSID data has several shortcomings, which makes it inappropriate for addressing many relevant questions. Moreover, a degree of caution is in order when interpreting some of the statistics based on this data.[13]

1. **Time Aggregation.** The PSID is well suited to study mobility at an annual frequency although more detailed job and employer questions (for up to two main and two secondary jobs) have been introduced since 1984. This makes it impossible to study the worker mobility patterns at a higher frequency. Monthly CPS data has a clear advantage in this respect.

2. **Invariant 1970 Occupation and Industry Classification.** The PSID uses the same 1970 Census of Population occupation and industry codes throughout the 1968-2001 period. On the one hand, having the same coding system throughout the whole period is beneficial since it allows one to consistently measure the patterns in occupational and industry mobility. On the other hand, it is clear that some of the occupations people worked in in the early 1990s were not even in existence when the 1970 Census classification was developed. Kambourov and Manovskii (2008) discuss that when new occupations appear, workers in those occupations will be coded as belonging to the "not elsewhere classified" occupational categories of the outdated classification. This implies that, over time, these "not elsewhere classified" occupations themselves represent collections of new occupations. Kambourov and Manovskii (2008) report that the fraction of workers employed in the "not elsewhere classified" occupational categories in the PSID increases from 14% to 21% over the 1968-97 period. Starting from the 2003 interview year, the PSID switched to coding occupations using the 2000 Census classification.

3. **Representativeness.** A concern with any longitudinal study is whether sample attrition makes the sample less representative of the population over time. Multiple

[13] Our main focus in the paper has been on discussing the (March) CPS data. Our discussion of the PSID data is more cursory; more detailed analysis can be found in Kambourov and Manovskii (2008, 2009a,b).

studies, summarized in Brown, Duncan, and Stafford (1996), show that the attrition in the PSID is small and random and that the PSID remains representative of the US non-immigrant population. However, those who have immigrated to the United States since 1968, as well as the children of such immigrants, are not part of the regular PSID sample. Thus, the PSID becomes less representative of the total population over time as such immigrants become a larger share of the population. To address this concern, a refresher sample of post-1968 immigrant families and their adult children was added to the PSID in 1997. While Fitzgerald, Gottschalk, and Moffitt (1998) found that the effects of the PSID non-representativeness due to immigration are small for a number of variables including demographic characteristics, income, and labor force status, we do not know what effects it might have on measures of worker mobility. The difficulty in answering this question lies in the fact that the CPS does not ask about the date when individuals might have immigrated to the US. The Census asks this question, but is only a cross-section making it impossible to compare mobility rates.

4. **Noisy Employer Data.** Brown and Light (1992) conducted a detailed review of the reliability of responses to job tenure questions in the PSID, found a great deal of inconsistency, and proposed various ways for dealing with it. Identifying genuine employer transitions is important on its own, but also because this is essential for identifying genuine occupational switches after 1980 when the Retrospective Files are not available. Such procedures were developed in Kambourov and Manovskii (2009b) who also document the amount of errors in identifying the genuine occupational and industry switches that each of these procedures entail. A researcher must be aware of the magnitudes of these errors and choose the procedure which is most appropriate for the question at hand.

5. **Noisy Occupational Affiliations.** Similar to the CPS, there is a substantial error in identifying the genuine occupational affiliations of workers in the PSID, especially after 1981 when Retrospective Files are not available. Thus, while occupational switches can be identified fairly reliably by using this data, there is substantial uncertainty about the true identity of the occupation the individual works in. As in the CPS, the statistics on net mobility and on the differences in occupational wages are suspect on the PSID data as well and have to be interpreted with caution. There appears to be no data source in the US on which these statistics can be reliably estimated.

6 Conclusion

Measuring the properties of worker mobility across occupations, industries, and jobs, as well as understanding its causes and consequences, is at the forefront of current research in macro and labor economics. One of the primary sources of data that the profession relies on in this research program is the Current Population Survey and its annual March supplement. The main reasons for the widespread use of this source of data are the large sample size of the CPS survey, the long period over which it is collected, and its perceived high quality. The Current Population Survey is the source of the US official labor market statistics and serves as a model for other household surveys, both in the United States and in many other countries.

Despite the prominence of this survey in measuring various labor market stocks and flows, the literature has largely ignored several of the important features of this data which are the consequences of (changes in) interview design, post-interview coding of data, imputation of the missing data, and the properties of non-response rates both over time and across population subgroups. In this paper, we argue that taking these features of the data into account changes the interpretation of some of the prominent findings in the literature.

In particular, we have contrasted the (March) CPS with the PSID Retrospective Occupation-Industry Supplemental Data Files as sources of occupation and industry affiliation data for measuring the dynamics of worker mobility in the United States. While we presented all the evidence based on the measures of occupational mobility, the evidence based on industry mobility is very similar.

Due to the adopted coding methodology, the PSID Retrospective Files minimize the coding error in identifying occupational and industry mobility. The coding methodology employed by the CPS differs and, as a consequence, the (March) CPS data is characterized by a substantial amount of noise when it comes to identifying occupational and/or industry switches. Researchers should be aware of this problem and of the potential effects on the estimates it may have.

Moreover, we find that the March CPS data provides a poor measure of annual occupational mobility. Instead, it likely measures mobility over a two- to three-month period. This is also important for researchers studying worker reallocation across occupations, firms, or industries using that dataset, in particular for calibrating and estimating structural models where the timing assumptions are crucial (e.g., Kambourov and Manovskii (2009a)).

We also found that changes in imputation and data processing procedures have a sizable effect on measures of worker reallocation. In particular, taking them into account changes some of the important conclusions in the literature.

Finally, we discussed that while the PSID Retrospective Files minimize the coding error in identifying occupational and industry mobility, they have a number of shortcomings such as a severe time aggregation, an invariant occupational classification, non-representativeness of the immigrant population, and noisiness in identifying genuine occupational affiliations, especially after 1981. In the end, the appropriateness of a particular data set depends on the question a researcher is interested in. The PSID and the CPS have unique advantages and disadvantages. In this paper, we outlined the key issues researchers should be aware of when analyzing worker mobility using this data and interpreting the findings.

References

Artuc, E., S. Chaudhuri, and J. McLaren (2007): "Trade Shocks and Labor Adjustment: A Structural Empirical Approach," Working Paper 13465, National Bureau of Economic Research.

Bollinger, C. R., and B. T. Hirsch (2006): "Match Bias from Earnings Imputation in the Current Population Survey: The Case of Imperfect Matching," *Journal of Labor Economics*, 24(3), 483–519.

Brown, C., G. J. Duncan, and F. P. Stafford (1996): "Data Watch: The Panel Study of Income Dynamics," *Journal of Economic Perspectives*, 10(2), 155–168.

Brown, J. N., and A. Light (1992): "Interpreting Panel Data on Job Tenure," *Journal of Labor Economics*, 10(3), 219–257.

Davis, S. J. (2008): "The Decline of Job Loss and Why It Matters," *American Economic Review: Papers and Proceedings*, 98(2), 263–267.

Dickens, W. T., and L. F. Katz (1987): "Inter-Industry Wage Differences and Theories of Wage Determination," Working Paper 2271, National Bureau of Economic Research.

Duncan, G. J., and D. H. Hill (1989): "Assessing the Quality of Household Panel Survey Data: The Case of the PSID," *Journal of Business and Economic Statistics*, 7(4), 441–451.

Faberman, J. (2008): "Job Flows, Jobless Recoveries, and the Great Moderation," Working Paper 08-11, Federal Reserve Bank of Philadelphia.

Fallick, B. (1993): "The Industrial Mobility of Displaced Workers," *Journal of Labor Economics*, 11(2), 302–323.

Feng, S. (2001): "The Longitudinal Matching of Current Population Surveys: A Proposed Algorithm," *Journal of Economic and Social Measurement*, 27(1/2), 71–91.

Fitzgerald, J., P. Gottschalk, and R. Moffitt (1998): "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics," Technical Working Paper 0220, National Bureau of Economic Research.

Heckman, J. J., and P. A. LaFontain (2006): "Bias-Corrected Estimates of GED Returns," *Journal of Labor Economics*, 24(3), 661–700.

Hirsch, B. T., and E. J. Schumacher (2004): "Match Bias in Wage Gap Estimates Due to Earnings Imputation," *Journal of Labor Economics*, 22(3), 689–722.

Jovanovic, B. (1979): "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87(5), 972–990.

Jovanovic, B., and R. Moffitt (1990): "An Estimate of a Sectoral Model of Labor Mobility," *Journal of Political Economy*, 98(4), 827–852.

Kambourov, G., and I. Manovskii (2008): "Rising Occupational and Industry Mobility in the United States: 1968-1997," *International Economic Review*, 49(1), 41–79.

——— (2009a): "Occupational Mobility and Wage Inequality," *Review of Economic Studies*, 76(2), 731–759.

——— (2009b): "Occupational Specificity of Human Capital," *International Economic Review*, 50(1), 63–115.

Krueger, A. B., and L. H. Summers (1988): "Efficiency Wages and the Inter-Industry Wage Structure," *Econometrica*, 56(2), 259–293.

Lee, D. (2005): "An Estimable Dynamic General Equilibrium Model of Work, Schooling, and Occupational Choice," *International Economic Review*, 46(1), 1–34.

Lee, D., and K. I. Wolpin (2006): "Intersectoral Labor Mobility and the Growth of the Service Sector," *Econometrica*, 74(1), 1–46.

Lillard, L., J. P. Smith, and F. Welch (1986): "What Do We Really Know about Wages? The Importance of Nonreporting and Census Imputation," *Journal of Political Economy*, 94(3), 489–506.

Liu, R., and D. Trefler (2008): "Much Ado about Nothing: American Jobs and the Rise of Service Outsourcing to China and India," Working Paper 14061, National Bureau of Economic Research.

Lucas, R. J., and E. Prescott (1974): "Equilibrium Search and Unemployment," *Journal of Economic Theory*, 7, 188–209.

Madrian, B., and L. Lefgren (2000): "An Approach to Longitudinally Matching Current Population Surveys (CPS) Respondents," *Journal of Economic and Social Measurement*, 26(1), 31–62.

Mathiowetz, N. A. (1992): "Errors in Reports of Occupations," *Public Opinion Quarterly*, 56(3), 352–355.

McCall, B. P. (1990): "Occupational Matching: A Test of Sorts," *Journal of Political Economy*, 98(1), 45–69.

Mellow, W., and H. Sider (1983): "Accuracy of Response in Labor Market Surveys: Evidence and Implications," *Journal of Labor Economics*, 1(4), 331–344.

Moscarini, G., and K. Thomsson (2007): "Occupational and Job Mobility in the US," *Scandinavian Journal of Economics*, 109(4), 807–836.

Moscarini, G., and F. Vella (2002): "Aggregate Worker Reallocation and Occupational Mobility in the United States: 1971-2000," Working Paper 02/18, The Institute for Fiscal Studies.

——— (2003): "Aggregate Worker Reallocation and Occupational Mobility in the United States: 1976-2000," Manuscript, Yale University.

Murphy, K. M., and R. H. Topel (1987): "The Evolution of Unemployment in the United States: 1968-1985," in *NBER Macroeconomics Annual*, ed. by S. Fischer. The MIT Press.

Oh, H. L., and F. J. Scheuren (1980): "Estimating the Variance Impact of Missing CPS Income Data," *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 408–415.

Polivka, A. E., and J. M. Rothgeb (1993): "Overhauling the Current Population Survey: Redesigning the Questionnaire," *Monthly Labor Review*, 116(9), 10–28.

PSID (1999): *A Panel Study of Income Dynamics: 1968-1980 Retrospective Occupation-Industry Files Documentation (Release 1)* Survey Research Center, Institute for Social Research, The University of Michigan, Ann Arbor, Michigan.

Rothgeb, J. M., and S. R. Cohany (1992): "The Revised CPS Questionnaire: Differences between the Current and the Proposed Questionnaires," *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 649–654.

Stewart, J. (2002): "Recent Trends in Job Stability and Job Security: Evidence from the March CPS," Working Paper 356, Bureau of Labor Statistics.

U.S. Bureau of the Census (2002): "Current Population Survey - Design and Methodology," Technical paper 63RV.

U.S. Bureau of the Census (1976): "Money Income and Poverty Status of Families and Persons in the United States: 1975 and 1974 Revisions (Advance Report)," in *Current Population Reports, Series P-60, No. 103*. U.S. Government Printing Office, Washington, D.C.

Table A-1: Matching Rates, Monthly CPS, 1976-2004.

Year	December-March		January-March		February-March		March-March	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
1976			93.0	56.0	94.9	57.0		
1977	92.0	57.0	93.0	56.8	95.0	57.2	77.8	54.5
1978	45.0	0	86.9	96.2	90.1	96.5	46.3	0
1979	85.0	95.0	86.8	95.9	90.1	96.3	73.5	93.6
1980	75.0	95.0	88.4	95.9	91.0	96.5	66.1	93.4
1981	86.0	98.0	88.7	98.7	91.1	98.8	75.7	97.7
1982	86.7	98.5	88.5	98.6	91.2	98.9	77.7	97.9
1983	87.3	98.4	88.8	98.6	91.4	98.8	76.6	97.8
1984	86.0	98.5	87.9	98.6	90.7	98.8	75.9	97.3
1985	86.3	98.2	87.3	98.2	90.6	98.5	73.7	97.4
1986	85.9	97.9	87.5	98.2	90.4	98.5	50.4	0
1987	85.3	98.1	88.0	98.4	90.2	98.5	74.7	97.1
1988	85.6	97.6	88.1	97.7	90.5	98.1	75.7	97.0
1989	85.8	97.7	88.4	97.8	90.8	98.2	76.4	96.6
1990	84.3	97.6	87.1	98.1	89.8	98.3	72.1	96.8
1991	86.5	97.9	87.9	98.0	90.6	98.3	76.0	96.7
1992	86.3	97.9	88.6	98.1	91.0	98.3	76.4	97.0
1993	86.5	98.1	88.5	98.3	91.5	98.3	76.9	97.1
1994	85.4	97.5	89.6	98.8	92.6	99.3	77.0	96.9
1995	87.6	98.1	92.5	84.4	94.6	84.5	74.1	97.1
1996	84.6	97.7	89.4	99.0	92.5	99.5	54.4	0
1997	87.9	98.5	90.6	99.1	92.7	99.5	77.8	97.3
1998	86.5	98.3	90.7	99.0	92.9	99.5	78.4	97.3
1999	86.8	98.0	90.3	98.9	92.5	99.4	78.3	97.2
2000	85.0	98.3	89.5	99.0	92.1	99.3	76.7	97.3
2001	86.5	98.2	89.7	98.8	92.7	99.4	78.7	97.3
2002	86.9	98.2	90.0	98.9	92.9	99.4	67.0	96.9
2003	86.8	95.3	90.5	98.9	92.8	99.3	78.6	94.5
2004	86.2	98.1	90.1	98.9	93.1	99.3	78.8	97.0

Notes: The matching rates in column (1) report the fraction of individuals who can be matched based on the formal identifiers provided in the monthly CPS. The matching rates in column (2), reported as a fraction of those already matched by the first method, exclude all those who have been matched but show a different gender, race, or age which is at most three years apart.
Source: Authors' calculations from the monthly CPS.