Yale University
Department of Economics

Yale Working Papers on Economic Applications and Policy
Yale University
P.O. Box 208268
New Haven, CT 06520-8268
DISCUSSION PAPER NO. 19

Occupational and Job Mobility in the US

Giuseppe Moscarini and Kaj Thomsson
Yale University

July 2006

Occupational and Job Mobility in the US

Giuseppe Moscarini
Yale University
giuseppe.moscarini@yale.edu

Kaj Thomsson
Yale University
kaj.thomsson@yale.edu

Abstract

We propose a new methodology to measure and to study worker mobility across occupations and jobs in US Census data at the monthly frequency. Our approach builds on two main ideas. First, we use the longitudinal dimension of matched monthly CPS files to evaluate each occupational transition in the context of the transitioning worker's employment history over four consecutive months. Second, we rely on the post-1994 Dependent Coding of occupations, and additional filters, to (in) validate potentially suspicious transitions.

When we apply our methodology to the 1979-2004 period, we obtain new estimates of the average levels and time series patterns of these labor market transitions. We find that about 3.5% of workers employed in two consecutive months report different 3-digit occupations. This flow is procyclical, mildly rising in the 1980s and falling after 1995, faster after the 2001 recession. Based on the results regarding occupational mobility, we can impute information to the numerous missing answers to the job-to-job (or Employer-to-Employer, EE) survey question. We revise upward current estimates of the average EE rate since 1994, to 3.2% per month. This rate mildly declines in 1994-1997, mildly rises in 1997-2000 and falls significantly and continuously in 2001-2004. This pattern suggests a very persistent negative impact of the latest two recessions on job-to-job mobility.

Keywords: Occupational mobility, Job mobility, Turnover

JEL Classification: E32, J62, J63

1. Introduction

In this paper, we propose a new methodology to measure and to study worker mobility across occupations and jobs in US Census data at the monthly frequency. We build on our findings on occupational mobility to shed new light on Employer-to-Employer (EE) transitions.

There are at least three reasons why economists are interested in occupational and job mobility. First, the concept of "mobility" (defined broadly) is tightly linked to the notion of "opportunity," which is a central part of the history of ideas and the formation of a national identity in the United States. In particular, the perception of great social mobility, measured among other things by income mobility and occupational mobility (between or within generations), is typically considered one of the key components of what is sometimes referred to as "American exceptionalism" (the idea that the US is different in some fundamental ways from most other countries).[1] Second, a prominent tradition in macroeconomics, going back at least to Schumpeter (1939), emphasizes the continuous reallocation of resources across heterogeneous production units as the "mode" of aggregate business fluctuations and economic growth. If capital is a quasi-fixed factor, technological progress can only be implemented through the "creative destruction" of installed capital and the reallocation of labor to new production processes. In this paper, we attempt to relate our work to this tradition by focusing on occupational mobility at a level of disaggregation that, in our view, corresponds well to a change of technology for a worker. That is, occupational mobility as measured here should be closely related to the creation and/or destruction of technology in the Schumpeterian sense. The same tradition has partly inspired modern equilibrium job search theory, which is empirically informed by several types of labor market flows. Lately, the interest has focused on job-to-job transition, and we revisit this flow in light of what we learn about occupational mobility. Third, on the labor supply side, human capital may be to a large extent occupation-specific, thus lost upon a transition.

The focus of this paper is on the actual *measurement* of occupational mobility. We believe that, when attempting to understand the nature of occupational mobility in the US, the Current Population Survey (CPS) is, for several reasons, a natural source of empirical information. First and foremost, since our focus is largely macroeconomic, we require a representative sample. The CPS is designed for this very purpose, and is superior in this

[1] For a discussion of the relationship between American exceptionalism and the concept of occupational mobility, see Ferrie (2005).

sense to any US longitudinal database (such as the National Longitudinal Survey of Youth or the Panel Study of Income Dynamics). Secondly, the CPS is the primary source of US official labor market statistics. Hence, comparisons with official reports and labor market information is greatly facilitated by the use of the CPS.

Measuring labor market transitions is a very difficult task in general, because mobility rates are typically small and very sensitive to measurement error in the labor market state of interest. This is particularly true in the case of occupations, which, unlike the three employment states (employed/unemployed/out of the labor force), number from dozens to thousands, depending on the level of disaggregation. The variety of tasks that US workers perform makes the description by the survey respondent often problematic, and small errors in stock classification lead to large spurious flows. In this difficult scenario, Murphy and Topel (1987) and Kambourov and Manovskii (2004) point out that there are particular problems with the annual (March) CPS files. Both studies provide convincing evidence of vast classification error in occupations, leading to hugely inflated transition rates, and the latter study notes the lack of a well-defined time-period over which the mobility that can be derived from the March CPS Files is measured.

We use *monthly* rather than annual CPS data in order to minimize problems of time aggregation and to exploit the longitudinal nature of monthly data to deal with occupational coding error. Our approach builds on two main ideas. First, since 1994 the survey includes a battery of Dependent Coding questions that are meant to minimize false transitions. Second, both before and after 1994, we inspect each single occupational transition between two consecutive months under the magnifying lens of a "global" view of that worker's employment history over four consecutive months, including one month before and one after the two months spanned by the transition. Sequences of four consecutive occupations that involve two transitions forth and back to the same initial occupation, and that do not correspond to changes in industry or class of workers or to active job search in the past month, are suspicious and should be, to a large extent, discarded. Similarly for other types of transitions that appear extremely rare in records that are (based on Dependent Coding) reliable, and extremely frequent otherwise.

Our main results are readily summarized. While raw data are extremely noisy before 1994, and still somewhat suspicious after 1994, applying the described filters makes pre- and post-1994 occupational mobility lower and almost identical on average over the entire 1979-

2004 period. About 3.5% of workers employed in two consecutive months report different 3-digit occupations, implying a significant change of career. This flow is procyclical, mildly rising in the 1980s and falling after 1995, especially after the 2001 recession.

Next, based on these findings, we then turn our attention to job-to-job (EE) transitions after 1994. Here, our contribution is different. Previous studies of this very important flow in the monthly CPS (most notably, Fallick and Fleischman (2004) and Nagypal (2004)) necessarily discard 12% of the sample who are missing an answer to the job change question. We are able to assign reliable information on occupational mobility to those records with missing job mobility. We then exploit the correlation between occupation and job mobility to impute an answer, when missing, to the job mobility question. This leads us to revise upward current estimates of the EE flow, to 3.2% per month. The time series pattern of the revised EE rate remains roughly in line with results from previous studies: flat in the 1990s, falling in 2001-2004. But a different interpretation is also possible, where the EE mobility rate mildly declines in 1994-1996, rises in 1997-2000 and drops precipitously from the 2001 recession to the end of our sample in March 2004. This second interpretation suggests a very persistent negative impact of recessions on job-to-job mobility.

Finally, we document that occupational and job mobility are far from being perfectly correlated. About 40% of occupational transitions are internal to an employer, presumably promotions, and over 1/3 of all job-to-job transitions involve no change whatsoever in the narrowly defined tasks that the worker performs. All of these flows show a coherent negative trend after 2001.

Although our exercise consists essentially of measurement, it is certainly not neutral nor theory-free. Our cleaning of raw data on labor market transitions exploits the unique high-frequency longitudinal nature of the monthly CPS to assess the plausibility of various kinds of transitions. We exploit this feature to extend pre-1994 the significant improvements introduced in the survey after that date, but we also inspect and further correct the much more reliable post-1994 observations. All this requires a fair amount of judgement calls. Standard theories of worker turnover underlie some of our choices. Turnover begets turnover, that is, the chance of a separation from a job or occupation is much higher if a separation occurred recently. But, at the same time, it appears unlikely that two consecutive monthly switches, bringing the worker back where he was, would occur very frequently, because turnover is in part motivated by dissatisfaction with the initial match.

We take confidence in our choices because of the consistency of our results, both across subperiods and with other surveys. After 1994, we rely on detailed questions about the actual change of tasks and activities to isolate the likely miscoded records, namely the independently coded occupations. In those records, we find an extent of measurement error that is almost exactly the same as before 1994, although we use different procedures to address those two periods. The final estimates we settle upon are also very similar over the two subperiods, when data were collected differently. Finally, the 3.5% average monthly mobility rate across 3-digit occupations is easily compatible, after taking into account correlated transitions and difference in definitions, with the annual 20% rate found by Kambourov and Manovskii (2005) in an extensive revision of PSID information from 1968-1997. The time series behavior of our and their series is similar in the overlapping 1979-1997 period.

Section 2 contains a brief introduction to the monthly (basic) CPS, Section 3 describes the procedure of matching cross-sectional information from different months, Section 4 discusses the main issues with the occupational information in the CPS, Section 5 overviews our approach to dealing with these issues, Section 6 applies this approach and produces our estimates of genuine occupational mobility, Section 7 addresses job-to-job mobility, Section 8 concludes.

2. The Current Population Survey

The Current Population Survey (CPS) is a monthly survey of about 50,000 households, that has been conducted by the Bureau of the Census for the Bureau of Labor Statistics for more than 50 years. However, because the information required to reliably match individuals over time is only available in later surveys, for our purposes information is available for the past 25 years only.[2]

Until very recently, studies of labor market transitions in the CPS have used the March Income Supplement files, which is a collection of files with data released on annual basis. Given our interest in labor market transition, in this study we use instead the Basic Monthly files. As we will see shortly, this high frequency allows us to fully exploit the limited, yet still very informative longitudinal component of the CPS, and to minimize attrition.

[2] Most of the overview information presented in this section is directly based on the official description of the CPS at the Bureau of Labor Statistics website (www.bls.census.gov/cps). We use a particular package−The CPS Utilities−consisting of data, documentation and Windows software developed by the company UNICON, INC in order to simplify the process of finding and extracting data from the CPS.

In fact, despite not being primarily intended for longitudinal analysis, the Current Population Survey contains a panel component and can be used to follow individuals over short periods of time. In each month the full CPS sample is divided into eight "rotation groups," with each housing unit being interviewed for four consecutive months, then removed from the sample for an eight month period and finally interviewed for another four months. Hence, in any month one-eighth of the sample households, the first rotation group, are interviewed for the first month, one eighth are interviewed for the second month, one eighth for the third month etc. Since the interviewers follow housing units (i.e. addresses) and not families or individuals, attrition can then occur for one of three main reasons: temporary absence (hospitalization, imprisonment, vacation), migration (to go to college, to enlist in the military, to form a family, to follow or to separate from a spouse, and for work-related reasons, including retirement), and mortality. To minimize attrition and the resulting possible sample selection, we will restrict attention to the first four months in sample. More details follow in the next section.

The CPS has several advantages and disadvantages over panel datasets, such as the PSID and the NLSY, for studying various measures of labor market states (employment/unemployment, even disaggregated by various demographics, occupation and industry membership etc.) and transitions (across employment states, occupations, industries etc.). The first advantage is the very large number of individuals in the sample. This means that only the CPS contains a fully representative sample of the US population, and that only the CPS can potentially eliminate most sampling error when measuring employment distributions across cells that number in the dozens or even many hundreds, such as 3-digit occupations, birth cohorts etc. The second advantage is the high frequency of observations and time series dimension, as the CPS is conducted monthly, as opposed to panels that conduct yearly interviews about the entire history of the previous 12 months. The monthly frequency minimizes (although does not completely eliminate) time aggregation problems due to multiple within-period undetected transitions, a severe problem with annual data. The third advantage is the wealth of information about demographics, income etc., which exceeds that of panel data.

The main disadvantage of the CPS for our purposes is its address-based nature, which aligns attrition with geographical mobility, potentially correlated with occupational mobility. In contrast, panel datasets continue to track the same individuals wherever they are, although panels too suffer from significant attrition because of their much lower interview frequency.

We address this issue in detail in the next section. Another disadvantage of the CPS is the very limited longitudinal dimension, as individuals are followed for eight (non consecutive) months, as opposed to decades for panel surveys. This is an unavoidable consequence of the much richer information set provided by the CPS: since so many questions are asked, they can be asked only so few times, lest becoming harassment. Our cleaning procedure is designed specifically to take this limitation of the CPS into account; four consecutive observations on the same individual is the least that we need to create our filter for spurious occupational transitions. Finally, coding of occupations and industries in the CPS changes with every decennial census, while panel datasets have a uniform coding maintained throughout the period.[3] We also address this issue in later sections.

3. Matching of Files

Measuring labor market transitions is a delicate business. Survey data are typically plagued by both sample attrition and moderate measurement error in the labor market state of interest (employed/unemployed, occupation, industry, class of worker etc.). Even when these bear no significant impact on measures of stocks, they can lead to significant bias in measured transition rates.

In principle, the re-interviewing process in the monthly CPS should allow us to match three-fourths of the sample in any given month to the next month, while one fourth of the sample exits due to rotation (though individuals in their fourth month can be linked eight months forward). The various kinds of attrition and absence mentioned earlier reduce the fraction of individuals that can actually be matched. In addition, recording errors in the variables that are meant to uniquely identify individuals further complicate the linking of individuals across months. Both failures of matching are a concern if and only if they are related in any way to occupational mobility.

In this sense, the three main sources of attrition−temporary absence, migration and mortality−appear to be unrelated to possible career mobility, hence unlikely to introduce any selection bias in the sample, with one important exception: migration for work-related reasons (either of the individual himself or of his spouse). In the appendix we derive an upper

[3] This is a mixed blessing because the Census revises coding every ten years precisely to avoid carrying outdated classifications. For example, suppose that a 3-digit occupational group grows in employment membership and in the variety of tasks it encompasses, so as to lead the Census Bureau to sub-divide it into two separate 3-digit codes. Then, the old coding system will lead to miss mobility among these new categories, thus to understate true mobility.

bound to occupational mobility by estimating annual geographical mobility from the March CPS files, converting it into an estimate of monthly geographical mobility, and adding the latter to the aggregate level of occupational mobility that we obtain from our sample. We show that this correction is minor. In addition, to minimize selection to begin with, we later focus *only on individuals in their first four months in sample*, to eliminate the additional attrition that takes place during the eight months between the two sampling periods. For both reasons, in the main part of this paper we proceed as if migrants were no different from non-migrants when it comes to changes in occupation. In this section, we illustrate in details these issues and the matching procedure.

3.1. Matching Strategy

The first thorough evaluation of matched CPS data is Peracchi and Welch (1995).[4] Their study provides an evaluation of the determinants of attrition from CPS, and the selected bias in the matched sample that may arise if the matched sample differs in a non-random way from the non-matched sample (or, equivalently, from the basic cross-sections). Two recent articles build on the work of Peracchi and Welsh and evaluate in depth the design of the matching criteria.

The first of these studies, Madrian and Lefgren (2000), is a very careful evaluation of the benefits and drawbacks associated with different matching criteria. Their key, starting observation is that, while in theory it should be possible to match individuals across different files/months using only the variables HHID (household identifier variable), HHNUM and LINENO (individual identifiers), in practice, due to recording error, this procedure will give rise to both falsely non-matched observations and falsely matched observations. Madrian and Lefgren evaluate the trade-off between these possible errors and propose a straight-forward validation algorithm: potential matches are considered valid if variables indicating sex, race or age agree, where "agreeing" means that sex and race have to be the same and the age variables in the two datasets have to be within an range of three years apart. In all other cases, the potential matches are discarded.

The second recent study, Feng (2001), adds marital status to the sex, age and race variables that are used by Madrian and Lefgren to assess the validity of potential matches. The algorithm suggested by Feng is somewhat more lenient, rejecting less potential matches

[4] Peracchi and Welch (1993) use a matching program described in an earlier paper by Welch (1993).

than suggested by Madrian and Lefgren.

In this study, we adopt the somewhat more conservative matching approach suggested by Madrian and Lefgren (2000).[5] That is, we first match datasets based on the formal identifier variables (HHID, HHNUM, LINENO), and then we evaluate the validity of potential match-pairs by comparing the sex, age and race variables. There is, however, a key difference between our study and these previous studies: our study is based on matching the basic monthly data instead of annual data. With the basic monthly files, instead of only two observations per individual, we have potentially eight monthly observations per individual. This is something we will return to in later sections, where we use these multiple observations to clean measured occupational mobility from spurious transitions changes. In addition, using monthly data minimizes the time period between two observations for a given individual. A priori, one would expect this to lead to lower attrition rates and a matched sample that is more similar to the original sample than in the case of a matching based on yearly data.

3.2. Results of Record Matching

Aggregate match statistics for men aged 16-64 during the years 1979-2004 are shown in Figures 3.1, 3.2 and 3.3. We show results only for individuals that are in the CPS for their second month. We make a distinction between the pre-1994 and post-1994 data, since in 1994 a large overhaul of the CPS aimed to increase the reliability of a number of variables. Generally, we expect the post-1994 data to be of higher quality than the pre-1994 data, which turns out to be the case. In the tables, the "naive match rate" is the share of individuals who, in their second or sixth month in the sample, could be matched with an observation one month *ahead* based on the identifier variables (HHID, HHNUM, LINENO and, in some cases, STATE) only. Individuals are matched if and only if all identifiers agree. The "valid match rate" gives the share of individuals who, in their second or sixth month in the sample, first were matched naively and then were validated by the criteria (described above) based on sex, race and age.

Because our focus is on occupational mobility, we restrict attention to individuals in the labor force, i.e. individuals who are employed, on temporary layoff or looking for employment. When computing occupational mobility, we will further restrict attention to men

[5] That is, we choose to be somewhat more conservative in our assessment of the validity of matches than Feng (2001). Simple checks indicate, however, that the difference in the final match rates between the two different proposed criteria would be small.

	1979-1993		1994-2004	
	Frequency	Percent	Frequency	Percent
Non-matched	48,397	4.74	18,950	3.17
Matched	972,714	95.26	578,297	96.83
Total	1,021,111	100	597,247	100

Figure 3.1: Naive Match Rates, Labor Force

	1979-1993		1994-2004	
	Frequency	Percent	Frequency	Percent
Non-matched	62,532	6.12	26,212	4.39
Matched	958,579	93.88	571,035	95.61
Total	1,021,111	100	597,247	100

Figure 3.2: Final Match Rates, Labor Force

	1979-1993		1994-2004	
	Frequency	Percent	Frequency	Percent
Non-matched	45,759	5.66	18,964	3.97
Matched	763,158	94.34	458,396	96.03
Total	808,917	100	477,360	100

Figure 3.3: Final Match Rates, Employed

who are employed in two consecutive months. Although this choice has obvious selection consequences, we make it for two reasons. First, it provides the correct measure of reallocation of existing employment across occupations. Second, the only alternative is to impute a "shadow occupation" to the unemployed, for example the last occupation held, which imposes non-mobility during the unemployment spell, while there are good reasons to believe that the unemployed are more prone than average to switch careers. Therefore, both methods underestimate occupational mobility, but our method (of focusing on the employed) less so.

Figures 3.1 and 3.2 display the naive and valid match results for the entire labor force, while Figure 3.3 displays the valid match rates when the sample is restricted to include only employed individuals. As expected, the match rates differ for the pre-1994 and post-1994 samples. To interpret the results, consider the match rates for the individuals in the entire labor force who are in their second month: for the post-1994 period, 96.83 percent of these observations are matched naively between their second and third month. Then, some of these naive/potential matches are considered invalid and therefore discarded, and the final result is that 95.61 percent of all individual in the labor force who are in their second month in the CPS are matched with an observation one month ahead. The corresponding result when we restrict the sample to only include employed individuals is that 96.03 percent are matched. Match rates are generally higher post-1994, a symptom of the improvement in data quality following the 1994 overhaul of the CPS.

3.3. Impact of Missing Matches on the Measurement of Labor Market Transitions

These match rates are considerably higher than the rates that have been achieved with the March CPS files, for instance by Madrian and Lefgren (2000). Still, about one out of 20 individuals in either their second or sixth month are lost when we attempt to match them with their own observation one month ahead. The implications for our interest in occupational mobility are not perfectly clear, since we do not know what happens to the individuals who fall out of the sample. It is clear, however, that the interpretation of this attrition depends upon the similarity between the full cross-sectional sample and the sample of individuals that we are able to match. This is the basic issue raised by Welch (1993), and analyzed in Peracchi and Welch (1994).[6]

[6] Somewhat surprisingly, Peracchi and Welch (1994) conclude that while selection on the matched records can introduce various biases, "no major bias appears in the estimates of transitions between labor force

Characteristic	Labor Force All	Labor Force Matched	Labor Force Non-Matched	Employed All	Employed Matched	Employed Non-Matched
Age						
Average Age	37.4	37.5	36.2	37.7	37.8	36.7
Gender						
Male (%)	N/A	N/A	N/A	N/A	N/A	N/A
Female (%)	N/A	N/A	N/A	N/A	N/A	N/A
Race						
White (%)	88.0	88.3	85.2	88.7	88.9	86.5
Black (%)	8.0	7.8	9.8	7.4	7.3	8.9
Other (%)	4.0	3.9	4.8	3.9	3.8	4.6
Veteran Status						
Veterans (%)	27.7	27.5	29.4	28.1	27.9	30.4
Marital Status						
Married (%)	64.5	65.3	58.3	66.2	66.8	61.1
Education						
< High School (%)	17.8	17.7	22.9	16.6	16.3	20.4
High School (%)	57.8	57.9	56.8	58.0	58.0	57.6
College (%)	24.3	24.7	20.3	25.4	25.7	22.0
Total Number of Obs.	1,574,330	1,411,235	163,095	1,472,492	1,327,048	145,444

Figure 3.4: Sample Characteristics, Month-in-Sample=2

Figure 3.4 displays the observable characteristics of individuals in their second month. There seems to be only one truly significant difference: the individuals in the matched sample have more education than the average for the full sample. In addition, there might be a selection on unobservable variables that is significant for the case of occupational mobility.

Given our focus on occupational mobility, we need to quantify the relative importance of the possible causes of failed matches, in particular of attrition/absence vs. erroneous personal identifiers. Our demanding criteria in terms of person identifiers and demographics are aimed to eliminate the possibility that different individuals erroneously match, and to minimize the impact on the measured frequency of occupational transitions. As we will see later, on average 96.5% of the individuals in the employed sample are occupational stayers month-to-month, and 3.5% are movers. Consider now the following two scenarios. In one, we discard, due to age miscoding, individuals who should be matched. Suppose these are 1% of the sample. This subtracts at most 1%·3.5%=0.035% of the employed sample from the numerator of the mobility ratio (movers/sample), and 1% of the sample from the

states after controlling for sex, age and labor force status at the time of the first survey." Note that this is a conclusion regarding the potential problems in the measurement of a flow variable (transitions in and out of the labor force), and not just about the measurement of the underlying stock variable.

denominator. Because this ratio is close to 0 to begin with, these two omissions in the numerator and denominator would not change the ratio significantly. This would remain true even if the individuals with miscoded ages were two or three times more likely than average to change occupation.

In the second scenario, we match individuals who should not be matched according to age. Suppose these are 1% of the sample. Since these are truly different individuals, almost all of them will appear as occupational movers at the 3-digit level. This procedure adds 1% of the sample to both the numerator and denominator of the mobility ratio, raising it roughly from 3.5% to 4.5% of the sample, a fairly significant impact. Therefore, we assume that the first scenario applies, and that the average 5% non-match rate is due exclusively to attrition, and not to erroneous identifiers. This is the most conservative choice, which can lead to an only negligible mismeasurement of occupational mobility.

Although 5% attrition at the monthly frequency may seem large, in real life there are many reasons for temporary or permanent attrition. Fortunately, only one of them, geographical mobility for job-related reasons, is relevant to occupational and job mobility. All the other reasons—such as geographical mobility due to retirement and college, as well as incarceration and hospitalization—disqualify the individual from the labor force anyway, so for our purposes they play the same role as in a proper longitudinal dataset, and can be safely ignored.

Because we do not know what happens to individuals who leave their address even if temporarily, we have to use out-of-sample information on the US population. The appendix shows how we can use information on retrospective geographical mobility from the annual March CPS files. In March of each year, respondents are asked whether they moved in the last year. We estimate, from annual data after allowing for correlated moves, that the fraction of employed workers who change address every month is less than 1% and very mildly cyclical. Therefore, no more than 1/5 of failed matches can be attributed to geographical attrition, the rest being other types of attrition that do not affect the occupational mobility of the employed sample. In addition, there is no reason to believe that all of these individuals change occupation. In fact, more educated people are much less likely to change occupation and much more likely to change location, often exactly because they do not want to change occupation. Therefore, it is highly implausible that more than 0.5% of the labor force change location *and* occupation each month, and the exact fraction is likely to be even lower. Since

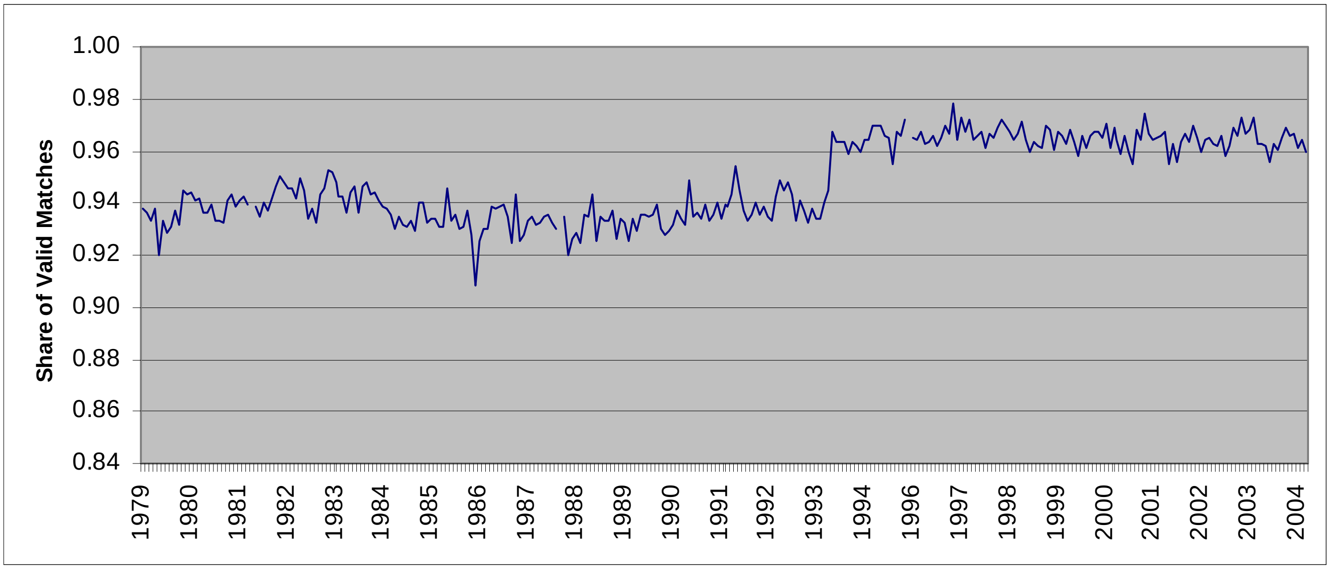

Figure 3.5: Match Rates, Employed

our estimates of monthly occupational mobility for matched individuals always exceed 3% and are strongly cyclical, we conclude that geographical attrition in the monthly CPS cannot have a large impact on our estimated level and cyclicality of true occupational mobility. The effect can be more significant, and must be kept in mind, because geographic movers are likely to change jobs.

We also notice a new fact. In Figure 3.5 we plot the match rate over time. It is clear that the match rate is mildly countercyclical, especially in the aftermath of the 1982 and 1990 troughs. That is, it is easier to match files in a recession. This is suggestive of the role played by migration for job search reasons, because labor market mobility is typically procyclical, as is occupational mobility. However, as said, geographic attrition, which is the only real reason of concern, is not only very small relative to measured occupational mobility, but also almost acyclical.

Finally, it should be noted that between some months the individual identifier was, for various reasons such as the confidentiality of the respondents, changed and/or non-unique. Hence, these months cannot be matched. This is the case for pre-1978 data, and for some months in 1985, 1994 and 2004. In order to correct for this, we simply removed observations from the months that cannot be matched, so all the results presented below are derived only for the months in which the individual identifier variables at least theoretically should uniquely identify individuals during their (maximally) eight months in the sample.

4. Occupational Coding Issues

In this paper, an "occupation" is a code defined by the US Census 3-digit occupational classification system. Examples are bartenders, accountants, computer programmers, chemistry teachers and automobile mechanics. The 3-digit level of classification was chosen primarily for two reasons: it best captures the notion that we believe most people have about an "occupation", and it is the level of classification that corresponds most closely to the skills that different jobs require. That is, moving between two different jobs with different 3-digit occupational classifications is likely to imply a change in skills that are necessary to carry out the job. Therefore, measuring occupations at this level corresponds most closely to the notion of "labor technology," with labor input being differentiated by the tasks involved and the kind of training required.[7]

Although the 3-digit Census codes correspond nicely to concepts of theoretical interests, using the occupational classification in the CPS to measure aggregate worker mobility is not without problems. In particular, the system of classification has changed several times. This complicates any analysis using time-series or longitudinal data. First of all, during the period we consider (1979-2004), the occupational codes were changed on three occasions, at the beginning of 1983, 1992 and 2003, following each decennial Census. Second, the interviewing technique changed dramatically in 1994, when the technique called "Dependent Coding" was introduced. Third, occupational data for some individuals are missing and have been imputed, with the obvious implication that occupational mobility might be overestimated. We now discuss the first two issues, whereas the third issue is addressed after the results are presented.

4.1. Census Occupational Classification

The changes in occupational codes in 1992 and 2003 were negligible, involving a handful of categories eliminated or added, and a very small fraction of employment. The 1983 change was significant, as the number of 3-digit occupations increased by about 17%, from 375 to 438,

[7] The alternative to using 3-digit level would be the 2-digit occupation coding system, which we consider too coarse. For instance, both "Chefs" and "Waiters" (3-digit occupations) fall under the 2-digit code of "Food Preparation and Related," and both "Architects" and "Biomedical Engineers" fall under the same 2-digit code of "Architecture and Engineering Occupations." Clearly, the 3-digit codes here identify what most people think of as an occupation, whereas the 2-digit codes are groupings of related but distinct occupations.

which per se would tend to increase measured mobility. This is a straightforward consequence of the methods of classification, and cannot be adjusted without introducing subjective assessments on the way in which different classification systems should appropriately be compared. We have made no attempt to correct for these changes, so the reader should be aware of the expected changes in mobility that are simply due to changes in codes and not fundamental changes in the economy. That is, the levels of occupational mobility pre- and post-1983 are not directly comparable. However, by removing all the matched month-pairs that span two different classification systems, we do make sure that no occupational mobility observation is due to the fact that the same job, for the same individual, is classified differently in two different months.

4.2. Dependent and Independent Coding

In order to understand the fundamental change in the reliability of the occupational mobility data before and after 1994, we briefly describe the interviewing procedure.[8] Before 1994, occupations were coded into numbers independently (**Independent Coding**). That is, *every* month, respondents were asked:

1. For whom did you work?
2. What kind of business is this?
3. What kind of work were you doing?
4. What were your most important activities or duties?
5. Were you
    - An employee of a private company, or business, or individual for wage or salary commission
    - A federal government employee
    - A local government employee
    - Self-employed in own business, professional practise, or farm
    - Working without pay in a family business

[8] This description is based on Polivka and Rothgeb (1993) and the UNICON documentation reporting the original questions from the Census protocols.

- Never worked

and this information was later used by CPS staff to assign occupation (and industry) codes to the individual.

This pre-1994 interviewing process had at least two serious problems. First, asking these questions was very cumbersome for the interviewer, and respondents typically complained about answering the same questions repeatedly. Secondly, and more importantly for this paper, asking these questions independently from month to month introduced a significant amount of spurious shifts in occupation and industry. The detailed nature of the information in the 3-digit occupation codes (as well as the 3-digit industry codes) implied that even if an individual answered the questions listed above in essentially the same way in two different month, the reviewers frequently assigned different 3-digit codes for the two months.

Indeed, in a small validation study of occupational coding based on company records and employees' descriptions of their own tasks, Mathiowetz (1992) finds a roughly 50% error when occupations are coded without telling the coders that the two records concern the same individual. More remarkably, when told that the two records did come from the same individual, expert coders still found a 12% disagreement rates between the company record and the employee's description of the employee's task, although each coder knows that they should agree.

To reduce the interview burden and the possibility of misclassification, a number of changes were introduced in 1994. First of all, "Dependent Coding" (sometimes referred to as "Dependent Interviewing") was introduced. Individuals interviewed in successive months were asked the following questions:

1. "SAMEJOB." Last month, it was reported that (name/you) worked for (input company name). (Do/Does) (you/he/she) still work for (input company name) at (your/his/her) MAIN job?
   - Yes (ask next question)
   - No, Don't know, Refused (skip to independent industry/occupation questions)
2. "CHDUTY." Have the usual activities and duties of (your/his/her) job changed since last month?
   - Yes (skip to independent industry/occupation questions)

- No, Don't know, Refused (ask next question)

3. "SAMEACT." Last month (name/you) (was/were) reported as (a/an) (input occupation) and (your/his/her) usual activities were (input duties 1) (input duties 2). Is this an accurate description of (your/his/her) current job?
   - Yes, Don't know, Refused (end series and use Dependent Coding)
   - No (skip to independent industry/occupation questions)

and if the Dependent Coding was used, then the same occupation as the previous month was automatically assigned. So the pre-1994 "Independent Coding" questions were asked after 1994 only when the interviewer could be confident that there had been a genuine change of activity. In addition, when Independent Coding was used after 1994, a direct question - "That is, what is your occupation?" - was added. Therefore, we refer to pre-1994 as "Unconditional Independent Coding", and to post-1994 as "Conditional Independent Coding" of different occupations, where in the latter case the coding is conditioned on received information on an ascertained change of employer and/or activity and task. Evidently, the large miscoding reported by Mathiowetz (1992) applies only to the former, pre-1994 Unconditional Independent Coding.[9]

Of all men matched and employed in months in sample 2 and 3, about 5% on average change occupation each month after 1994 (see Figure 6.1 in the next section). This is the result of raw data. Before 1994, raw occupational mobility is a whopping 33.8%. We now assess the quality of occupational coding post-1994, and address later the more problematic pre-1994 period.

Dependent coding reduces spuriously measured shifts in occupation status, as confirmed by a small validation study presented in Polivka and Rothgeb (1993). In a validation sample of 1,392 individuals, they compare (i) occupational mobility measured with pre-1994 techniques and (ii) occupational mobility measured with post-1994/dependent techniques, with

[9] From Mathiowetz (1992)'s study, Kambourov and Manovskii (2004) conclude that Dependent Coding in the CPS is problematic for the purpose of measuring occupational mobility, because 50% of (say) 9% independently coded occupations in the validation study sums up to a very large potentially spurious flow of 4.5%. But the 50% miscoding refers to Unconditional Independent Coding, so to the pre-1994 period. In fact, the coders in Mathiowetz (1992)'s experiment did not have access to the three additional Dependent Coding questions. Even the 12% discrepancy found between description of what is known by the coder to be the same job should not concern us, because Dependent Coding is precisely meant to avoid that the respondent describe his/her tasks when they have not changed. The evidence that we present below shows that post-1994 Conditional Independent Coding is fairly reliable.

(iii) the results from a detailed study of occupational mobility carried out by a consulting firm, WESTAT Inc., on a subsample of 406 individuals. The estimated aggregate mobility with post-1994 techniques (7 percent) fell within the bounds as identified by WESTAT (5.9-7.4 percent), while the mobility measured with pre-1994 techniques was much greater (39 percent). This pattern fits well with the non-cleaned mobility measures for the years before and after 1994.

We rely on post-1994 data to provide a benchmark measure of occupational transitions, that we believe to be accurate for two main reasons. First, the main goal of Dependent Coding is precisely to make sure that occupations are coded independently *if and only if* the occupation has in fact changed with very high probability. To this purpose, four different questions are asked, three of which do not require any description by the respondent, and two of which specifically ask whether activities have changed, even reminding the worker of the past answer. Even if the last month's description was inaccurate, presumably the respondent should be able to reconstruct whether we are talking about the same position as a month ago. For those who do change occupation, it is irrelevant whether they provide accurate descriptions, all that matters is that the occupation has in fact changed. That is, following an ascertained true change of activity, the accuracy of the independent descriptions is irrelevant for our purposes. On the other hand, inaccurate responses are very unlikely to produce a false negative (no mobility, when in fact the occupation did change). The validation results in Polivka and Rothgeb (1993) provide hard evidence for these arguments: the independently coded question implied a true mobility in the acceptable range.

These considerations still leave the door open for some miscoding post-1994. Consistently with this idea that not all independently coded records are equal, we provide additional evidence of our own. The main purpose of this paper is to design a "cleaning" algorithm of spurious occupational transitions, that builds on the limited longitudinal dimension of the data to replicate and to refine the logic of Dependent Coding and to be applied pre-1994, as well as an additional consistency check with a change of industry, class of workers or job search activity. The next section illustrates our two procedures and shows that, once applied to post-1994 records, they eliminate most suspicious independently coded transitions, leaving us with a fairly homogenous and clean sample. For this reason, we feel confident that our two cleaning procedures combined produce a fairly clean post-1994 sample. We now go into the details of our cleaning procedures.

5. Identifying Occupational Mobility with Matched CPS Data

In addition to exploiting Dependent Coding after 1994, our approach to distinguish valid from spurious changes in occupational status has two parts. First, we use the structure of each individual's sequence of (maximally four) consecutively reported occupations to identify sequences that appear "implausible." For instance, an individual who is moving from one occupation to another and then directly back to the first in consecutive months−say, waiter-waiter-bartender-waiter−is considered a potentially "invalid mover." Second, we use other variables whose difference over the two months is likely to be correlated with changes in occupation, in order to validate possible occupational mobility.

5.1. Exploiting the Longitudinal Dimension: The FLAG Filter

Our main filter exploits the limited but still rich longitudinal dimension of the monthly CPS. For this filter to apply, we focus on a transition between two months t and $t+1$ such that we have potentially also information at least about months $t-1$ and $t+2$. This allows us to identify the "trajectory" of three consecutive occupational transitions, i.e. to analyze them jointly rather than in isolation. This implies that we need information about at least four consecutive months, which is exactly what the CPS contains. Although we could also use information on months in sample 5-8, we focus on the first four months to minimize sample attrition and the resulting selection.

In order to use each individuals sequence of (four) consecutive observations on occupations, we have to identify all different possible cases. We are mainly after misclassified mobility, since the problem due to miscoded non-mobility is likely to be minor. From now on, MOB refers to a dummy variable indicating whether two occupation codes for the same individual differ in months-in-sample 2 and 3 (which would be a perfect measure of occupational shifts in the case of no misclassifications).

For records with an occupational transition, MOB=1, we now introduce a variable FLAG, which takes on different values to indicate the type of transition. To illustrate, assume that A, B, C and D represent different 3-digit occupations, and N stands for the case where occupation information is not available. Consider the possible sequences of codes in the first four months in the sample, for an individual who has valid and different codes in months 2 and 3 (a condition that defines MOB=1). The first sequence, A-A-B-B, is the "ideal case". As another example, the sequence A-A-B-C is plausible, based on the typical decline in the hazard rate of separation with tenure (Farber (1994)) as predicted by the canonical theory of worker turnover (Jovanovic (1979).) The first transition from A to B in the third month reveals a likely mismatch in occupation A and triggers a job-shopping process, which may not be immediately successful, and may lead the worker to keep trying again the next month with occupation C. For the same reason, the sequence A-A-B-A is implausible (although of course not impossible).

The FLAG variable, in almost all observations, takes on values from 1 to 13, as detailed in Figure 6.2. Any other possible sequence omitted in that table and in subsequent tables applies to less than 0.1% of the total sample. The fact that some of these occupation sequences are more likely than others to represent true occupational mobility forms the basis of our approach for cleaning the mobility data. The distribution of FLAG for the post-1994 transitions, cleaning according to the ANY3 (described below) and Dependent Coding filters provides us with a benchmark to evaluate the plausibility of a sequence, thus to decide whether to eliminate that kind of transition or not before 1994 or even after 1994 when the Dependent Coding questions have missing or suspicious answers.

5.2. Additional Checks: The ANY3 Filter

As mentioned earlier, an additional check is offered by variables that are likely to be correlated with the occupation (OCC) variable. If these variables change between two months for which the OCC codes change, the change in OCC codes is likely to represent true occupational mobility. We have identified the following variables, that are available both before and after 1994 and possibly correlated with changes in occupation:

- CLASS: class of worker (private firm, federal, state or local government, or self-employed)
- IND: 3-digit industry code
- LK/LKWK: has looked for work in past 4 weeks
- HOURS: hours of work
- WKSTAT: full-time or part-time work status
- FAMINC: family income

From the correlation between these variables and MOB, we found that in the post-1994 period (in which we believe the OCC/MOB data to be closer to the truth) changes in CLASS and IND happen basically only when MOB=1, and the correlation between LK/LKWK and MOB is also very strong (though of course not perfect). Conversely, changes in HOURS, WKSTAT and FAMINC correlates poorly with MOB. In the pre-1994 data these correlations maintain the same ranking but are all lower, as we would expect since the pre-1994 occupation data is of lower quality. In the text below, "ANY3" refers to the criterion which preserves an observed transition with a suspicious flag if *any* of CLASS, IND and LK/LKWK has also changed. Conversely, we consider spurious any change of occupation that corresponds to no change of industry and class of worker and does not follow active job search.

6. Results

6.1. Occupational Mobility After 1994

We now tackle the potential miscoding problem that might remain after 1994 due to Conditionally Independent Coding of some occupations. We define a record "Suspicious" if any of the following three events is true: either the answer to the first Dependent Coding question "same job?" is Blank, or the answer to any of the subsequent two questions (conditional on the series arriving there with previous valid Yes/No answers), "change of duty?" and "same activity?", is Blank.[10] All other records are "Valid." Our logic is as follows. A Blank answer to the first question triggers Independent Coding for unknown reasons, i.e., it is effectively Unconditional Independent Coding as pre-1994, so it is suspicious. A No answer signals a different job, so a change of occupation is plausible. Of those who reply to have the same job (Yes to the first question), we are suspicious only of the Blank answers to the second question, because those who say that they have kept the same job but changed duty should be expected to have changed occupation, so the independently coded occupations should be different with very high probability. Same for the third question: suspicious are those who have a Blank answer to the "same activity?' question after having reported the same job and no change in duties.

In Figure 6.1 we break down the sample in Valid and Suspicious records. Although Suspicious records are only 24,696/445,595=6% of the total, their occupational mobility is

[10] Here Blank means also "Don't know" or "Refusal," but the latter two categories are quantitatively negligible.

Occupational		All Records raw	Dependent Coding Answers: Valid	Suspicious raw	Suspicious any3-adjusted
Stayers	Frequency	422,632	408,675	13,957	17,587
	Percent	94.8	97.1	56.5	71.2
Movers	Frequency	22,963	12,224	10,739	7,109
	Percent	5.2	2.9	43.5	28.8
Total	Frequency	445,595	420,899	24,696	24,696
	Percent	100	100	100	100

Figure 6.1: Classification of 1994-2004 Data

43.5%, relative to 2.9% of the Valid records. This evidence corroborates our prior beliefs that Suspicious records are truly suspicious and Valid records are truly valid from the point of view of measuring occupational mobility. The mobility of Valid records is evidently due to Independent Coding of occupations that are almost surely different because of the valid answers to the Dependent Coding questions. This is the whole point of Dependent Coding, and as such it must be exploited and cannot be ignored. The Suspicious records are few in number but contribute a large amount to mobility, so we still have to deal with them.

We first pass post-1994 Suspicious transitions through our ANY3 filter. As shown in Figure 6.1, this cuts the proportion of Suspicious movers by a third, from 43.5% to 28.8%, already a significant improvement. Next, we pass the remaining Suspicious transitions through the "FLAG" filter. The benchmark is the flag distribution of the Valid records, which we assume to be immune from miscoding. Figure 6.2 reports the flag distribution for Valid and Suspicious occupational movers. We underline the frequencies of Suspicious Flags that show a great discrepancy from the frequencies in the Valid case. Based on this evidence, all Suspicious transitions with Flags 3 and 10, 11, 12 and 13 are eliminated and treated as spurious, that is, MOB is reset from 1 to 0 in those cases. Our decision is based on the fact that we do expect Suspicious records to be somewhat different from the Valid ones (there must be some reason why they are Suspicious), but we want to discard the cases where the discrepancy is very large. So we keep Flags 8 and 9, although different frequencies appear, and discard 3 and 10-13, because the discrepancy is quite large. While discarding all of Flag-3 Suspicious transitions may seem excessive (still almost 7% of Valid transitions have Flag 3), so is keeping all of Flag 8 and 9 Suspicious transitions, which have a frequency well over twice than among Valid transitions. While this filter calls for a subjective judgement, it

FLAG Number	Type	Valid Freq.	Valid Percent	Suspicious: Not cleaned Freq.	Suspicious: Not cleaned Percent	Suspicious: Cleaned by ANY3 only Freq.	Suspicious: Cleaned by ANY3 only Percent	All: Valid, cleaned by ANY3 & flags Freq.	All: Valid, cleaned by ANY3 & flags Percent
1	A-A-B-B	6786	55.51	823	15.97	492	6.92	7,278	46.39
2	N-A-B-B	1059	8.66	752	8.9	495	6.96	1,554	9.91
3	A-A-B-N	785	6.42	3,335	13.26	2,130	29.96	785	5.00
4	A-A-B-A	344	2.81	434	6.7	317	4.46	661	4.21
5	A-B-A-A	373	3.05	644	5.47	464	6.53	837	5.33
6	A-A-B-C	795	6.50	489	13.9	324	4.56	1,119	7.13
7	A-B-C-C	1063	8.70	1,039	9.58	719	10.11	1,782	11.36
8	A-B-C-D	323	2.64	761	8.85	461	6.48	784	5.00
9	N-A-B-C	163	1.33	287	5.2	193	2.71	356	2.27
10	A-B-C-N	204	1.67	737	3.61	534	7.51	204	1.30
11	N-A-B-N	237	1.94	1100	5.89	752	10.58	237	1.51
12	N-A-B-A	53	0.43	120	1.72	79	1.11	53	0.34
13	A-B-A-N	39	0.32	218	0.95	149	2.10	39	0.25
Total occup movers		12,224	100	13,046	100	7,109	100	15,689	100

Figure 6.2: Distribution of FLAG 1994-2004, by Dependent Coding Status

is clear from the table that one would arrive at roughly the same number of final transitions with different reasonable criteria.

The result of applying the ANY3 and FLAG filters after 1994, shown in Figure 6.3, is that we retain only 3,465 of the original 10,739 Suspicious transitions, and correspondingly cut the mobility rate of the 24,694 Suspicious records from 43.5% (Figure 6.1) to 14% (Figure 6.3). Importantly, this nearly 29.5% reduction in the measured mobility rate of Suspicious records *almost exactly* corresponds to our estimate of pre-1994 Independent Coding error for monthly CPS data, illustrated in the next subsection.

The final tabulation of cleaned transitions post-1994 is in Figure 6.3. Overall mobility is 3.52% per month. This is a weighted average of 2.90% of Valid transitions and 14.03% of Suspicious transitions cleaned by our ANY3 and FLAG filters. Given the relatively small number of initial and, even more so, of surviving Spurious transitions, we are confident that a very accurate estimate of average monthly occupational mobility after 1994 is in the 3.3%-4% range, and we maintain 3.52% as our preferred point estimate.

6.2. Occupational Mobility Before 1994

Before 1994, all occupations were coded independently. In our jargon, all of these records, whose total number is 714,086, are to be treated as Suspicious. Figure 6.4 illustrates the distribution of Flags for pre-1994 data. As mentioned earlier, raw occupational mobility is 33.8%, which is evidently too high at a monthly frequency and confirms our skepticism

Occupational		All Cleaned	Dep Coding Answers Valid	Suspicious cleaned
Stayers	Frequency	429,906	408,675	21,231
	Percent	96.48	97.10	85.97
Movers	Frequency	15,689	12,224	3,465
	Percent	3.52	2.90	14.03
Total	Frequency	445,595	420,899	24,696
	Percent	100	100	100

Figure 6.3: Final Occupational Mobility 1994-2004, by Dependent Coding Answer

about pre-1994 observations. However, this immediately tells us that the miscoding error in the CPS cannot be as high as 50% as reported in Mathiowetz (1992), but is likely close to a relatively more manageable 30%.

Our ANY3 filter eliminates almost 64% of all transitions, leaving us with a mobility rate of 12.2%. Next, we apply our FLAG filter by comparing the distribution of flags for the ANY3-cleaned pre-1994 data (Figure 6.4) to our benchmark, which is now the distribution of post-1994 cleaned records (last column of Figure 6.2.) This leads us to *retain* before 1994 *only* transitions associated to Flags 1,2,3, 6 and 7. Although this selection may appear drastic, notice that the retained Flags account for 75% of all transitions in the post-1994 cleaned sample, and nonetheless too many transitions of types (Flags) 6 and 7 survive before 1994. We also tried to either retain FLAG=4,5,8,9 or 10 transitions as valid if and only if they satisfied the ANY3 requirement, or to use the ANY3 criterion directly, without using the sequence/FLAG structure at all. In both cases, final mobility was above 15% per month, unrealistically high.

The final tabulation of cleaned pre-1994 occupational transitions is in Figure 6.5. The final mobility rate is 3.69% per month, a drop of about 30 percentage points from 33.8% in the raw (uncleaned) data. Final mobility is very close to the 3.52% we found in post-1994 data, and the effect of our filters is virtually identical when applied to pre-1994 transitions and to post-1994 spurious transitions: in both cases, mobility falls by roughly 30%. Since the new interview method was introduced in January 1994, we cannot compute mobility in December 1993 (between that month and January 1994). But, if much residual error remained pre-1994, we would expect final (cleaned) mobility to drop significantly between November 1993 and January 1994. We find the opposite: mobility rises from 4.4% to 4.9%. To get a sense of seasonal factors, the same two numbers for November 1992 and January

FLAG Number	Type	Raw Freq.	Raw Percent	Final (only ANY3 cleaning applied) Freq.	Final (only ANY3 cleaning applied) Percent	Final (ANY3 & flag cleaning applied) Freq.	Final (ANY3 & flag cleaning applied) Percent
1	A-A-B-B	18,815	7.8	6,520	7.5	6,520	24.8
2	N-A-B-B	7,150	3.0	2,798	3.2	2,798	10.6
3	A-A-B-N	7,093	2.9	3,171	3.6	3,171	12.0
4	A-A-B-A	32,535	13.5	10,085	11.6		
5	A-B-A-A	34,094	14.1	10,610	12.2		
6	A-A-B-C	16,869	7.0	6,532	7.5	6,532	24.8
7	A-B-C-C	19,361	8.0	7,301	8.4	7,301	27.7
8	A-B-C-D	74,956	31.1	26,867	30.8		
9	N-A-B-C	6,938	2.9	3,116	3.6		
10	A-B-C-N	8,178	3.4	4,013	4.6		
11	N-A-B-N	4,592	1.9	2,317	2.7		
12	N-A-B-A	5,413	2.2	1,838	2.1		
13	A-B-A-N	5,350	2.2	1,993	2.3		
Total occup movers		241,344	100	87,161	100	26,322	100

Figure 6.4: Distribution of FLAG 1979-1994, Before and After ANY3 and FLAG Cleaning

Occupational	Frequency	Percent
Stayers	687,764	96.31
Movers	26,322	3.69
Total	714,086	100

Figure 6.5: Final Occ. Mobility 1979-1994

1993 are 4.5% and 5.3% respectively, and for November 1994 and January 1995 they are 3.8% and 4%. So, there is always a seasonal increase in occupational mobility when comparing the months before and after the winter holidays, and the increase in 1993-1994 shows nothing special or suspicious. Finally, there is no reason to believe (with one exception, explained below) that any pre-1994 residual measurement error have any cyclical or trend component. We now turn to the time series properties.

6.3. Time Series Patterns of Occupational Mobility

We can now plot the time series of aggregate occupational mobility one month forward for all male individuals aged 16-64 who were in their second month of the interview sequence, matched and employed both in the second and third month in sample.

Some caveats are in order. As mentioned earlier, data throughout the entire 1979-2004 period are not fully comparable for two main reasons. In 1983 there was a significant change

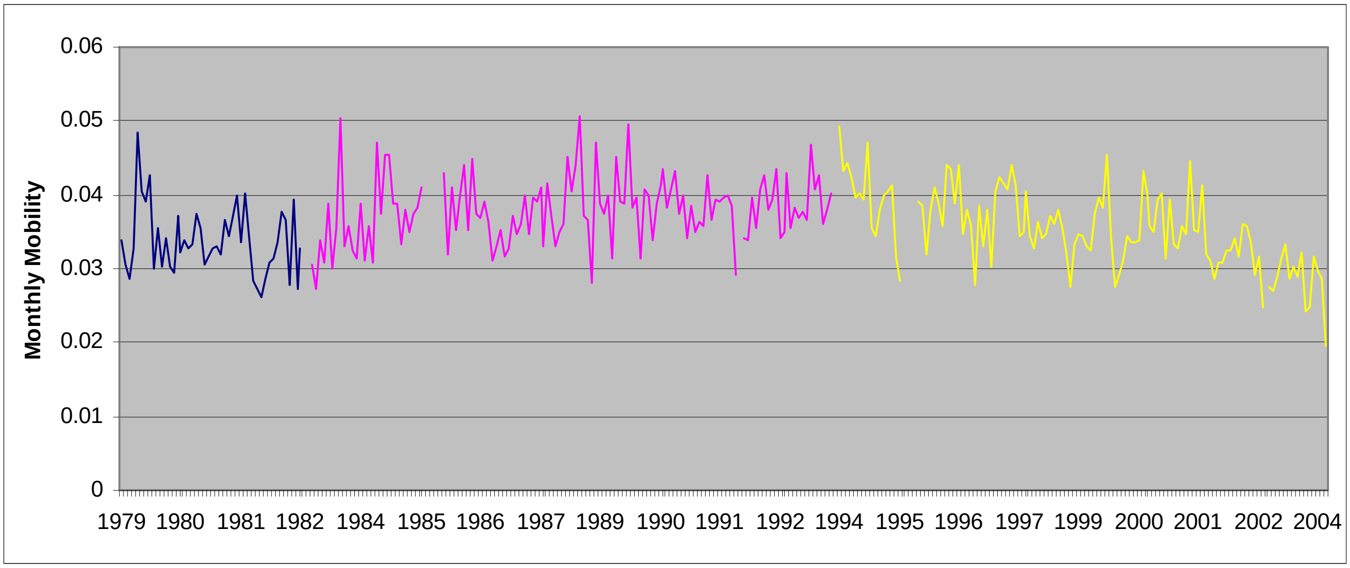

Figure 6.6: Occupational Mobility, Not Seasonally Adjusted

in the Census 3-digit occupational classification system, and in 1994 Dependent Coding was introduced. We cannot do much about the former, and we expect this change to raise measured mobility after 1983. We have tried to make pre-1994 data as similar to post-1994 observations as much as we can. At a minimum, we feel safe in analyzing the time series behavior of mobility in the following three subperiods separately: 1979-1982, 1983-1993, 1994-2004.

Figures 6.6 and 6.7 present the time series for the whole period and each subperiod, both the data in their actual form and after controlling for seasonality using X12A's Census methodology, with variable orthogonal seasonal factors and an AR(3) error structure. In the first subperiod, 1979-1982, there is a significant decline, which appears to be mostly a cyclical effect. As we see from entire 1979-2004 period, occupational mobility tends to be procyclical. In the second subperiod, 1983-1993, there is a definite initial rebound from the 1982 recession, followed by a long period of relatively constant mobility, with a slight dip in the 1989-1991 recession-recovery cycle. Perhaps most strikingly, after 1994, the period for which data are cleanest, we observe a definite and large decline in the pace of occupational mobility, with a clear dip in the 2001 recession that triggers a significant continued fall in 2002-2003.

Kambourov and Manovskii (2005) report similar time series of 3-digit occupational mo-

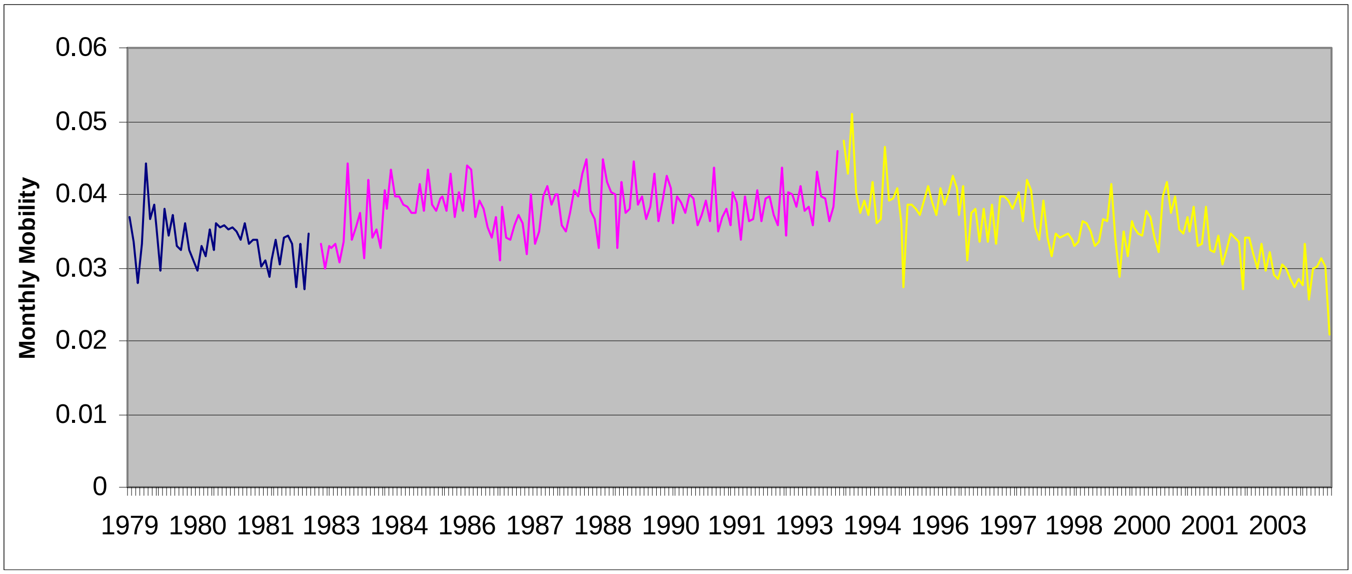

Figure 6.7: Occupational Mobility, Seasonally Adjusted

bility at annual frequency from the PSID in 1968-1997 and for a slightly different sample: men aged 23-61, heads of household, private sector, not self-employed. That is, they exclude (due to PSID data restrictions) non-heads and very young workers, as well workers aged 62-65, and also exclude government employees because they found that the occupational mobility of those workers plummeted, masking an overall positive trend. They include the unemployed and attribute them the last valid reported occupation from previous years. The average annual occupational mobility that they report from the PSID is about 20% at the 3-digit level. This is potentially consistent with our 3-4% at the monthly frequency when allowing for repeated and correlated within-year transitions. Indeed, if every individual changed occupation independently every month with chance 3.5%, then annual mobility would be around 35%, namely 1.75 times their reported rate. The well-known correlation in transitions (the hazard rate of separations from jobs is sharply decreasing in tenure) can easily explain part of the discrepancy from the annual mobility of 20% in the PSID as the result of time aggregation. In addition, Kambourov and Manovskii (2005)'s imputation of no mobility to the unemployed tends to bias down their estimate of overall mobility. Finally, in the PSID occupations are coded according to a unique, coherent, and by the end of the period somewhat obsolete, Census coding system, so it is likely to underestimate mobility as time goes by because of the lack of distinction between new occupations that emerged in the last three decades of the XX century (think software developers).

In terms of time series patterns, the PSID (Kambourov and Manovskii (2005)) shows a secular increase in 3-digit occupational mobility, more pronounced in the first half of the sample, before 1986, then in 1987-1997, although the annual data are fairly noisy, with frequent year-to-year variations of over $\pm 2\%$. The procyclical pattern is mild but visible, confirming our earlier finding. We rederived our time series using the same sample disposition of Kambourov and Manovskii (2005), again with the proviso of the three subperiods. The results with this sample were qualitatively unchanged from the findings derived with our baseline sample, as presented in Figures 6.6 and 6.7. Hence, our results agree with Kambourov and Manovskii's for the first few years in our sample period: we too find a mild increase in mobility between the late 1970's and the mid-1980s. However, there is a significant difference between our results and theirs (i.e. between the CPS and the PSID) when it comes to occupational mobility late in the sample, where we register a sharp decline, which continues after the end of their sample in 1997.

6.4. Imputed Occupations

Some values of the occupational codes are imputed, rather than assigned following either Dependent Coding or the current description, and are therefore likely to contain errors. The CPS Monthly files contain a variable, AOCC, that indicates whether the occupation code was imputed or not.[11] Imputation is likely to lead to miscoding and to overestimate occupational mobility. But discarding imputed records could lead to a non-representative sample, if some groups are overrepresented among the persons with imputed values. Clearly, this could be a potential issue with the results presented above.

Fortunately, this turns out to not be the case. In Figure 6.8 we tabulate the imputed and non-imputed records in our sample "cleaned" by the ANY3 and FLAGS procedures. The number of imputed occupations constitute a very small share of all observations with occupational changes, only rising to 11.15% after 1994. Although this last proportion is

[11] The CPS does allocations for two reasons: missing information, or inconsistent information. With regards to occupation data, the Census has certain requirements for consistency of the occupation, industry, and class of worker as a whole. Prior to 1984, no information about allocation is available. In 1984-1988, a binary allocation variable is set to one if either of these operations was used. After 1988, the Census expanded the range of allocation variable values to capture the many nuances of the allocation process. In general, values of 2-5 in the 1989-1993 time frame and values of 10-43 for 1994 to current are the same as the 1984-1988 variable value of one. We thank Gregory Weyland of the US Census Bureau for providing this information.

Occup.	1979-1983 Imputed	Not Imp.	1984-1988 Imputed	Not Imp.	1989-1993 Imputed	Not Imp.	1994-2004 Imputed	Not Imp.
Stayers	N/A	232,479	1,661	217,884	1,997	224,619	5,075	424,831
%		100	0.76	99.24	0.88	99.12	1.18	98.82
Movers	N/A	11,264	196	11,706	223	12,057	1749	13940
%		100	1.65	98.35	1.82	98.18	11.15	88.85
Total	N/A	243,743	1,857	229,590	2,220	236,676	6,824	438,771

Figure 6.8: Share of Imputed Occupation Observations

significant, it would not be wise to treat all of these transitions as spurious, and in fact our cleaning procedure should leave mostly imputed records that correspond to a genuine change of occupation, especially after 1994. Even counting all occupational transitions associated to an imputed record as spurious, an extremely conservative solution, overall mobility would not change at all before 1994, and would drop from 3.52% to 3.12% after 1994. Therefore, we feel much safer in keeping the remaining records with allocated values after 1994 in the sample as they are.

7. Job Mobility Since 1994

Another important type of labor market transitions is that from job to job. This is a flow of central interest to the calibration and estimation of equilibrium search models of the labor market. The monthly CPS after 1994 contains a SAMEJOB question which makes it ideally suited to measure Employer-to-Employer (EE) flows, although we still face time aggregation of multiple within-month EE transitions and short unemployment spells. This new source of information has been exploited recently by Fallick and Fleischman (2004) and Nagypal (2004). An EE transition occurs when the question SAMEJOB the following month receives a negative answer. They report a monthly mobility rate post-1994 of about 2.7% average. Neither of these studies report their treatment of Blank answers to the SAMEJOB question. This is a potentially serious issue, because as reported in Figure 7.1 these answers amount to 12% of our sample. Imputing all Blanks to either EE movers (Blank=No) or non-movers (Blank=Yes) makes the total EE rate range from 2.38% to 14.4%. Without additional information, however, it would be difficult to interpret this very large number of Blank answers to the key SAMEJOB question.

Occupational	Refuse	Don't Know	Blank	Yes	No	Total
Stayers	31	28	49,796	376,124	3,927	429,906
%	0.01	0.01	11.58	87.49	0.91	100
Movers	43	38	3,876	5,058	6,674	15,689
%	0.27	0.24	24.71	32.24	42.53	100
Total	74	66	53,672	381.182	10,601	445,595
%	0.02	0.01	12.05	85.54	2.38	100

Figure 7.1: Answers to SAMEJOB Question, by (Non-Cleaned) Mobility Status

7.1. Interpretation of Missing Answers

Our analysis of occupational mobility provides some guidance to better interpret the meaning of a Blank answer to the first Dependent Coding question (SAMEJOB) in month 3, whether the individual works for the same company as last month (the base month in sample 2) or not. From Figure 7.1, occupational transitions follow 1/4 of the time a Blank answer to the SAMEJOB question, almost 1/3 of the time a positive answer, and the remaining 42% a No answer (employer/company definitely changed). The second group, occupational transitions within the same company, are internal promotions, and account for a third of all occupational transitions, a new interesting fact in itself.

Since Blanks answers to the SAMEJOB question are so numerous and important among occupational movers, they must be investigated further. From Figure 7.2, which is the same as Figure 7.1 but with % frequencies computed by column, cleaned occupational mobility conditional on a Blank answer to SAMEJOB in month 3 is about 7%. We also verified that it is fairly stable over the 1994-2004 period. Compare this rate to the 63% occupational mobility rate conditional on an explicit negative answer to the same SAMEJOB question, and to the 1.3% mobility rate for those who answer Yes. The “Refusal” and “Don’t know” answers are negligible. This leads us to interpret most Blank answers as meaning Yes, namely the individual stayed with the same company and therefore was promoted, because he has a different occupation.

However, there is a five-fold (7.22%/1.33%) difference in occupational mobility rates between Blank and Yes. So, some significant fraction of Blanks may actually be in reality a No, namely be associated to a change of employer. If we exclude Blank answers and count job-to-job transitions to be only the relative frequency of No vs. Yes answers to the SAMEJOB question, we obtain a job-to-job mobility rate of 2.7% per month, exactly the

Final Occupational	Refuse	Don't Know	Blank	Yes	No	Total
Stayers	31	28	49,796	376,124	3,927	429,908
%	41.89	42.42	92.78	98.67	37.06	96.48
Movers	43	38	3,876	5,058	6,674	15,687
%	58.11	57.58	7.22	1.33	62.94	3.52
Total	74	66	53,672	381,182	10,601	445,595
%	100	100	100	100	100	100

Figure 7.2: Mobility across Jobs and Occupations (Cleaned)

same as that reported by Fallick and Fleischman (2004) and Nagypal (2004) for all possible months in sample and the same dataset and period. Therefore, restricting attention to months in sample 2 and 3 introduces no additional noise in this measure, and those authors must have excluded Blanks altogether. Following their procedure, promotions account for a full 43% of all occupational transitions, and over 1/3 of all job movers (SAMEJOB=NO) keep the same occupation. But our ANY3 and FLAGS filters preserve many Blank SAMEJOB records as valid for occupational transition purposes, so excluding all of these records from the computation of job-to-job transitions appears inappropriate. Since Blank answers to the SAMEJOB question are associated to occupational mobility that exceeds both the sample average and that of job stayers, we conclude that the 2.7% job-to-job mobility rate reported in the literature by ignoring Blank answers is likely to underestimate somewhat the true value.

7.2. Results

Following the discussion above, we now allocate Blank answers to Yes and No according to their relative frequencies of cleaned occupational mobility. With this allocation procedure, the final rate of EE flows rises to 3.2% per month. This is the average of the 1994-2004 time-series that we present in Figure 7.3. This series is seasonally adjusted, but significant high frequency volatility remains, presumably originating from sampling error. Although our estimate of the average EE flow is larger than previous ones, we still find that the flow was essentially flat in 1994-2001, with an initial decline in 1994, followed by a barely noticeable increase in the late 1990s, and then a pronounced fall in the 2001 recession, which continues through the end of the sample.

Next, in Figure 7.4, we address the correlation between job and occupational mobility. This correlation is positive but much weaker than one would expect, because over 40% of

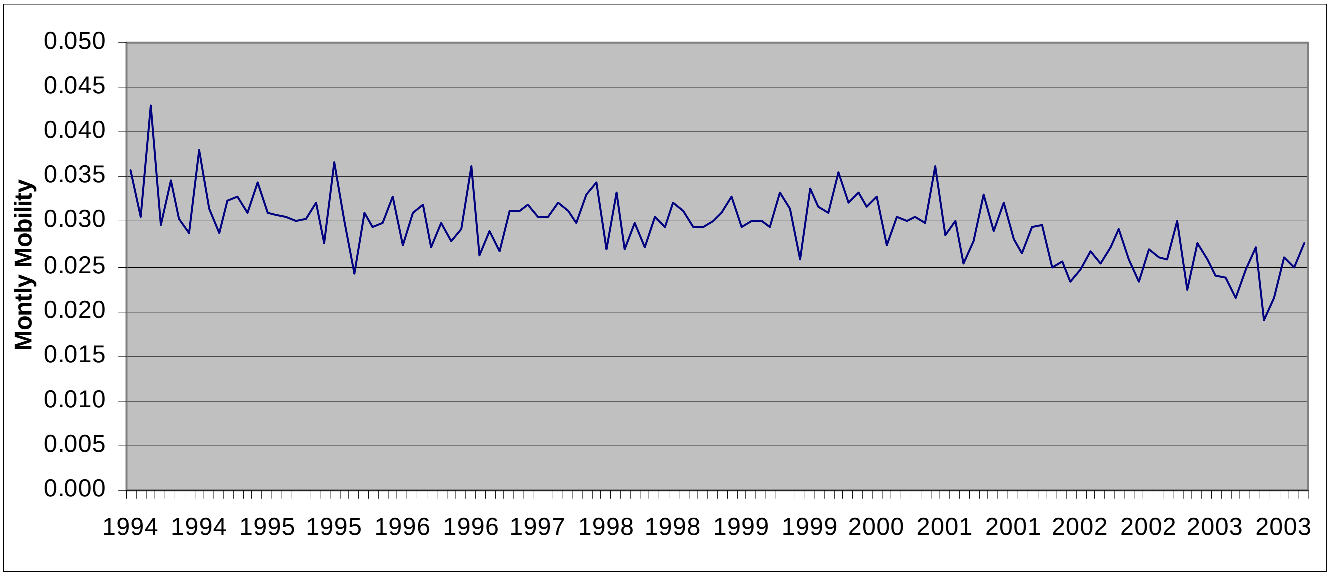

Figure 7.3: Job-to-Job Mobility 1994-2004, Seasonally Adjusted

Occupation\Job	Stayers	Movers	Total
Stayers	416927	4284	421211
%	95.55	0.98	96.53
Movers	6596	8546	15142
%	1.51	1.96	3.47
Total	423522	12830	436352
%	97.06	2.94	100.00

Figure 7.4: Job-to-Job and Occupational Mobility

occupational transitions involve no job change and over 32% of all job changes entail no change of occupation.

In Figure 7.5 we plot the 1994-2004 time series of disaggregations of job and occupational mobility. Since we are primarily interested in long-run trends and fluctuations at business cycle frequencies, and the seasonal adjustment makes almost no difference for this purpose, these series are presented unadjusted. The three series indicate, in order, the rate of joint job and occupational mobility, the rate of mobility across occupations while remaining with the same employer (presumably mostly "promotions,") and finally changes of employer without a change in 3-digit occupation, that we interpret as being transitions meant to pursue a "career." The sum of the first two is Occupational mobility in Figure 6.7, and the sum of the first and the third is Job-to-Job mobility in Figure 7.3.

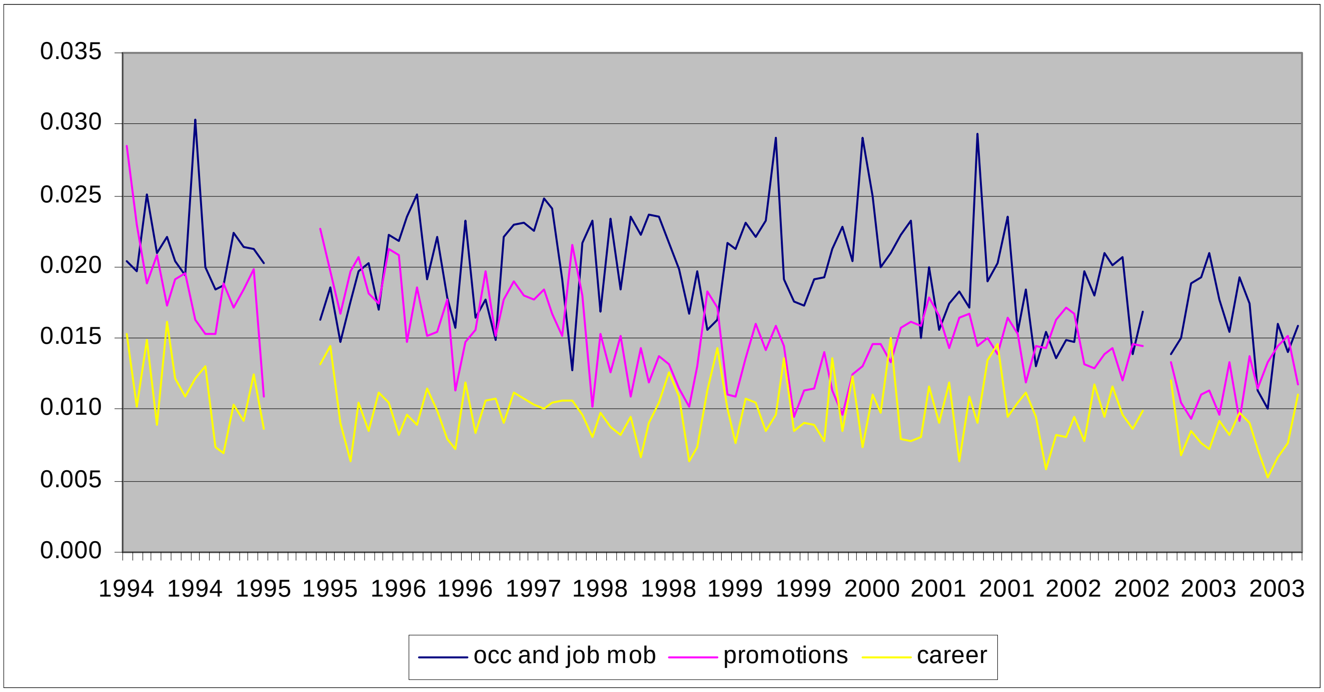

Figure 7.5: Disaggregated Measures of Job- and Occupational Mobility 1994-2004

In Figure 7.5, the levels of the three series are almost uniformly ranked in the order we listed them. Most transitions involve a simultaneous change of employer and tasks. Next come internal promotions, and last employer changes keeping the same occupation. However, all three flows are large. The smallest still hovers around 1% per month, about equivalent to the average layoff rate. Promotions are also fairly frequent. All three series experience a decline in 1994 and after the 2001 recession. The more interesting new fact is that, during the aggregate expansion of the 1990s, occupational switchers also increased their job mobility, and the relative incidence of internal promotions on occupational reallocation fell significantly. The aftermath of the 2001 recession brought back the two series roughly together as they were in 1994-1995, but at a much lower level.

8. Concluding Discussion

In this paper, we outline a new approach to measure and to study worker mobility across occupations and jobs in US Census data at the monthly frequency. The focus of the paper is on occupational mobility. In addition, we build on our findings on occupational mobility to shed additional light on employer-to-employer transitions. We exploit the limited longitudinal dimension of monthly CPS data, combined with the post-1994 Dependent Coding technique, to cleanse the very large spurious labor market flows that appear in the raw data. We find that occupational transitions at the 3-digit level (what most people think of as an "occupation") average at about 3.5% of employment per month. This is slightly larger than, and not so closely correlated with, our own revised measure of employer-to-employer transitions. Ocupational mobility is procyclical, and show a secular rise from the early 1980s to the early 1990s, and a secular decline from the mid 1990's on, accelerating towards the end of our sample in 2001-2004. Job-to-job mobility also sharply declines after 2001.

At this stage, we can only speculate on the sources of this recent prolonged decline in career mobility. Given the well-known negative correlation between age and any measure of labor market mobility, the aging of the US labor force is likely to have contributed to the decline. However, this aging of the population was a factor also in the 1980's, hence it cannot by itself account for the decline in mobility after the mid 1990's. An alternative explanation focuses on the effects of outsourcing of specific tasks of production to low-income countries. On the labor demand side, a reduction in the number of viable careers and the resulting concentration of the labor force into the service sector and more skilled occupations mechanically lead to a lower flow. On the labor supply side, an increase in the *perceived* uncertainty might make individuals more defensive in their career-shopping behavior, leading to a decrease in mobility. At this point, we are unable to assess the relative importance of these and other possible hypotheses, but we think it is an interesting direction for future research.

We have discussed throughout the paper the limitations of our approach. Here is one more, a somewhat more fundamental limitation. Because we are entirely focused on eliminating false transitions, we are not really interested in whether occupations have the correct code to begin with, but only that the numerical code assigned to an occupation changes if and only if the true underlying occupation does. That is, our cleaning procedure yields what we believe to be reliable estimates of transition rates, but not necessarily of the *stocks* of employed individuals in the different occupations. This prevents us, for example, from presenting reliable estimates of net employment reallocation across occupations. This is a topic of future research, and a goal still on our agenda.

Despite their limitations, we believe that the long, complete, and very high frequency estimates of labor market flows that we obtain from the premier and richest survey of the

US labor market may be useful to a variety of goals. Natural examples are the calibration of equilibrium search models to explain aggregate employment fluctuations, the structural estimation of search, sorting and turnover models, the estimation of the wage/mobility link in a regression context. In addition, from a microeconometric viewpoint, the wealth of information available in the CPS opens a wide range of possibilities.

References

Fallick, B. and C.A. Fleischman (2004), "Employer-to-employer flows in the US labor market: the complete picture of gross worker flows," Federal Reserve Board, Washington.

Farber, H. (1994), "The Analysis of Interfirm Worker Mobility ," Journal of Labor Economics, vol. 12(4), p. 554-593.

Feng, S. (2001), "The Longitudinal Matching of Current Population Surveys: A Proposed Algorithm," Journal of Economic and Social Measurement, vol. 21 (1/2), p. 71-91.

Ferrie, J. (2005), "History Lessons: The End of American Exceptionalism," *Journal of Economic Perspectives*, vol. 19(3), p. 199-215.

Kambourov, G. and I. Manovskii (2004), "A Cautionary Note on Using (March) CPS Data to Study Worker Mobility," Mimeo, University of Pennsylvania.

Kambourov, G. and I. Manovskii (2005), "Rising Occupational and Industry Mobility in the United States: 1968-1997." Mimeo, University of Pennsylvania and University of Toronto.

Jovanovic, B. (1979), "Job Matching and the Theory of Turnover." Journal of Political Economy, vol. 87(5), p.972-990.

Madrian, C. and L.J. Lefgren (2000), "An Approach to Longitudinally Matching Current Population Survey (CPS) Respondents." Journal of Economic and Social Measurement, 26, p. 31-62.

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

Murphy, K. and R. Topel (1987), "Unemployment, Risk and Earnings: Testing for Equalizing Wage Differentials in the Labor Market". In K. Lang and J. Leonard (Eds.), *Unemployment and the Structure of the Labor Market*, New York, Basil Blackwell.

Nagypal, Eva (2004), "Worker Reallocation over the Business Cycle: The Importance of Job-to-Job Transitions." Mimeo, Northwestern University.

Peracchi, F. and F. Welch (1994), "Trends in the Labor Force Transitions of Older Men and Women," Journal of Labor Economics, vol. 12, p. 210-242.

Polivka, A. E. and J. M. Rothgeb (1993), "Redesigning the CPS questionnaire," *Monthly Labor Review*, September, p. 10-28.

Schumpeter, J. (1939), *Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalist Process,* 2 Vols. New York and London, McGraw-Hill.

Welch, F. (1993), "Matching the Current Population Surveys," Stata Technical Bulletin, vol. 12, p. 7-11.

A. Appendix: An Upper Bound on Mobility

Although most individuals who could potentially be matched from one month to another are actually matched, a non-negligible share of potential matches are not actually matched in the final sample used in this paper. There exist a number of possible reasons, discussed in the Introduction, for the lack of matching of some individuals, i.e. reasons why individuals disappear from the sample between one month and the next. Since we focus only on individuals who are in the labor force for two consecutive months, some of the reasons - mortality, military enlistment, hospitalization and incarceration - are irrelevant. Even if data were available for these individuals for the second month, they would be removed from the sample. As of the other possible reasons for lack of matching, such as temporary absence due to vacation, there is no reason to expect that occupational mobility would systematically differ from the measured level of mobility of the matched sample. There is, however, one important exception: geographical mobility. For two reasons, we should worry that occupational mobility might be greater among the non-matched individuals. First, there very reason for the geographic move might be to start a new job, and possibly a new career (i.e. a change of occupation). Second, individuals that move for reasons that are not job-related (such as marriage) might not find a job in the same occupation at their new location, implying occupational mobility. Hence, by excluding all non-matched individuals from our sample, we are likely to underestimate occupational mobility, and some estimate of geographic mobility is necessary in order to get an idea of the potential downward bias in aggregate mobility.

Unfortunately, the monthly files in the CPS contain no information about geographic mobility. However, there is a question in the annual March CPS Supplement that asks individuals about migration within the past year. Our approach comprises the following four steps:

1. We first assume that our baseline value of occupational mobility, presented in the main text, provides a lower bound on mobility.
2. We use the March CPS to derive the annual migration rate of employed individuals in the US, for moves that occur across county lines.
3. The annual value is rescaled to a monthly migration rate.
4. Finally, we assume that all the people who move also change occupation, and add the geographic movers to the baseline value of occupational mobility. This delivers an upper bound on occupational mobility.

The estimated geographic mobility for employed individuals is displayed in Figure A.1. Note that, since we only have an annual variable for measuring geographic mobility, the estimated monthly values are obviously constant during one-year periods. Note also that around March 1985, the values is constant during a two-year period. This is because in 1985, individuals were asked about mobility between 1980 and 1985 instead of 1984 to 1985. Finally, note that we also use this - that mobility in 1985 refer to five years instead of one year - to adjust for the problem that some individuals may move more than once during the period that the question refers to.[12]

In order to derive an upper bound on occupational mobility, the monthly values of geographic mobility are simply added to the baseline values of occupational mobility. Figure A.2 contains these upper bounds. Note that since geographic movers not necessarily change their occupation, these upper bounds are likely to provide overestimations of occupational mobility.

[12] Specifically, we note that the value of five-year mobility (in 1985) is about three times greater than the following value of one-year mobility (in 1986), and not five times as large. We take this as a measure of repeated migration, and multiply all values of geopgraphic mobility with 5/3. If repeated moves are more likely, relative to all moves, on a five-year basis than on a one-year basis (which we expect to be the case), then this correction will overestimate mobility somewhat.

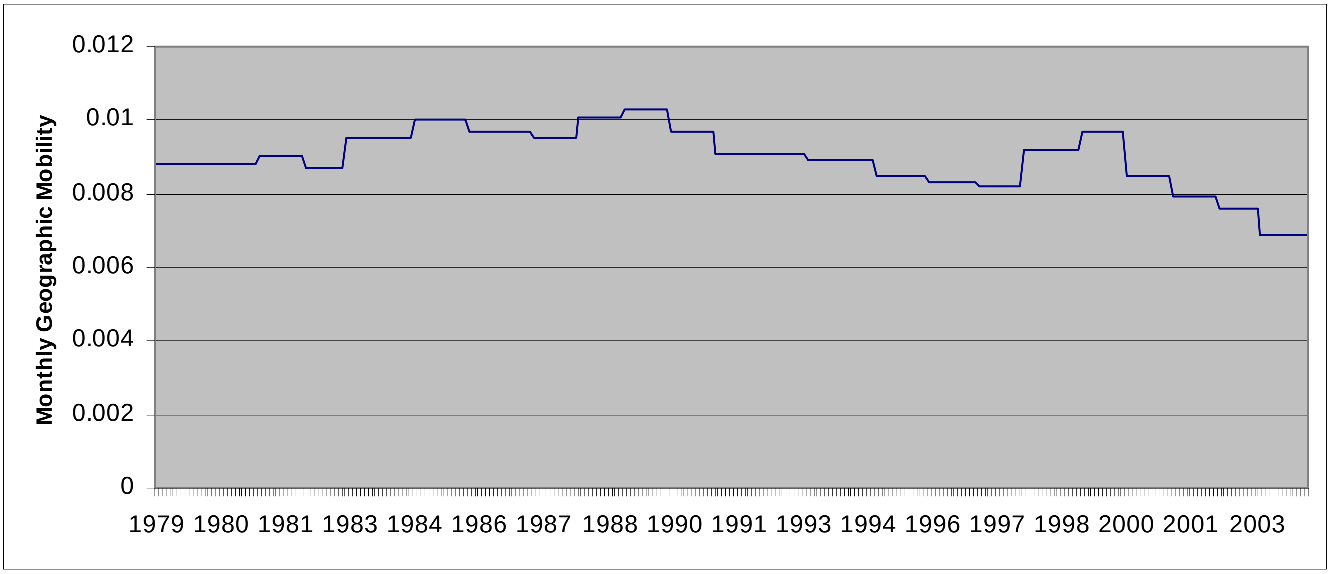

Figure A.1: Estimated Geographic Mobility

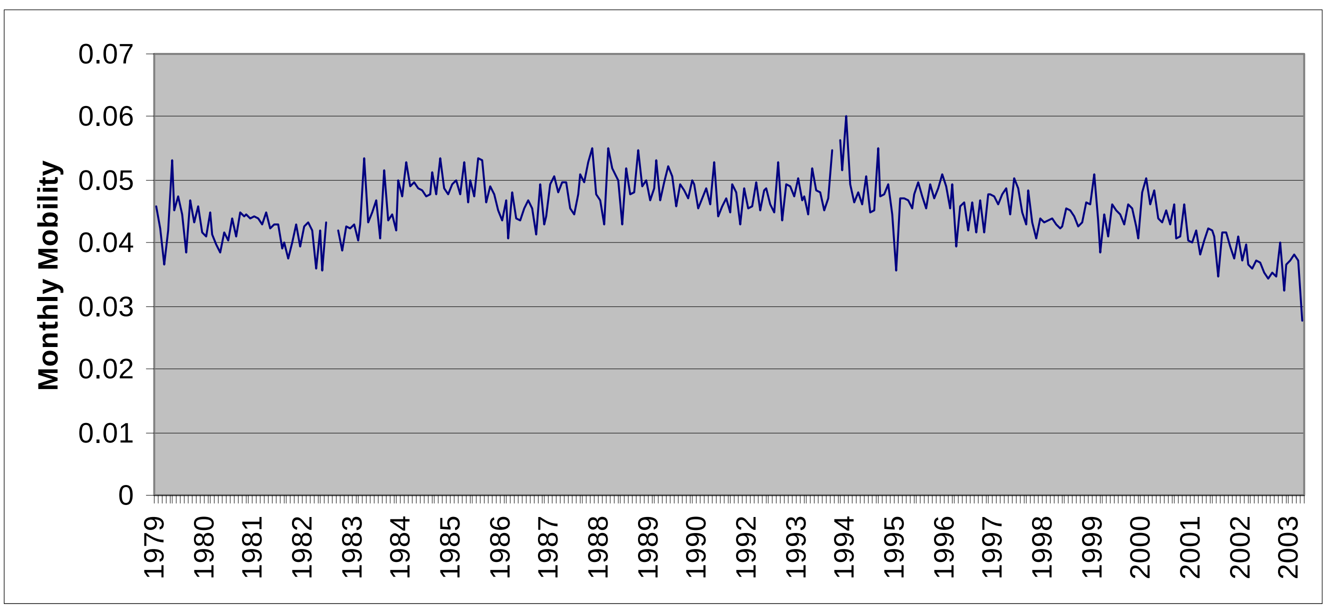

Figure A.2: Upper Bound on Occupational Mobility, Seasonally Adjusted