Title
Measuring U.S. labor market dynamics

Author
Nekarda, Christopher J.

Publication Date
2008

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Measuring U.S. Labor Market Dynamics

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Christopher J. Nekarda

Committee in charge:

Professor Garey Ramey, Chair
Professor James D. Hamilton
Professor Gordon H. Hanson
Professor David Phillips
Professor Valerie A. Ramey

2008

The dissertation of Christopher J. Nekarda is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2008

DEDICATION

I dedicate this work to my late step-father, Dr. Russell E. Ruth.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am indebted to my advisor, Professor Garey Ramey, for his support and guidance. I also deeply appreciate the input from the rest of my committee, Professors James Hamilton, Gordon Hanson, David Phillips, and Valerie Ramey. I am grateful to seminar participants at UCSD for their thoughtful comments and helpful critiques.

I acknowledge financial support from Valerie Ramey and the National Bureau of Economic Research through National Science Foundation grant SES-0617219.

I also thank the UCSD Institute for Applied Economics and Professor Marc-Andreas Muendler for providing essential computing resources.

VITA

2000	A. B. in Economics with highest honors, University of California, Berkeley
2003–2008	Graduate Teaching Assistant, University of California, San Diego
2008	Lecturer, University of California, San Diego
2008	Ph. D. in Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Measuring U.S. Labor Market Dynamics

by

Christopher J. Nekarda
Doctor of Philosophy in Economics

University of California San Diego, 2008

Professor Garey Ramey, Chair

This dissertation develops new data and methods for properly measuring U.S. labor market dynamics using large, nationally-representative household surveys. These data are used to assess potential biases arising from time aggregation and from geographic mobility.

Time aggregation is estimated using weekly labor force information from the Survey of Income and Program Participation (SIPP). The degree of time aggregation is large: gross flows estimated from monthly data understate the true number of transitions by 20 percent on average. However, time aggregation creates no meaningful cyclical bias in measured gross flows or hazard rates. Separation hazard rates calculated from the SIPP and the Current Population Survey (CPS) are strongly countercyclical and remain so after adjusting for time aggregation.

Using a new database that captures all longitudinal information in the CPS individuals who move can be identified. Comparing the behavior of the entire CPS sample with the subset known not to have moved provides a bound to the bias from geographic mobility. The cyclical bias from geographic mobility is small. At business cycle frequencies, the difference between the separation hazard rate calculated from the entire CPS sample and from a subset that are known not to have moved never exceeds 4 percent. There is little effect of mobility on the job finding hazard rate.

The weekly SIPP data identify direct employment-to-employment (EE) transitions. Abstracting from labor force participation, EE transitions account for one-half

of all separations from employment. Similar estimates using the CPS are twice as large however the CPS overstates EE transitions because of time aggregation. Separations to a new job are strongly procyclical while separations to unemployment are strongly countercyclical. The combination yields a nearly acyclical total separation rate. The weekly job finding rate is strongly procyclical.

Chapter 1

Introduction

This dissertation seeks to understand how labor markets evolve over time. Central to understanding labor market behavior is accurately measuring the dynamics of the labor market: the movement of individuals among employment, unemployment, and labor force nonparticipation. It develops data and methods for properly measuring U.S. labor market dynamics using large, nationally-representative data sets. These data are used to study potential biases arising from time aggregation and from geographic mobility.

The chapters that follow use data from two national surveys of U.S. households: the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP). The CPS, a monthly survey conducted to gather labor force information, is the canonical source for studying U.S. labor market dynamics. Because individuals in the CPS can be matched from one month to the next, researchers can identify labor force transitions by the change in a person's labor force status from one month to the next. However, two concerns have arisen about potential biases to dynamics measured by the CPS.

The first concern is that the monthly sampling frequency may be inappropriate to fully capture labor market dynamics. In particular, if labor force transitions occur more frequently than once a month, the month-over-month transition measure may combine multiple transitions into a single "aggregate" transition—hence, the term time aggregation. If these unmeasured spells occur with different propensity during a recession than during an expansion, the measured labor market dynamics may have a cyclical bias.

The second concern about the CPS involves geographic mobility. The CPS does not follow individuals that move away from addresses selected for interview. Because geographic mobility often occurs for economic reasons, particularly involving unemployment, there is concern that the dynamics captured by the CPS may be biased from sample attrition.

Properly assessing these concerns requires developing new data sources and a new approach to longitudinal research using CPS data. To study the bias from time aggregation, I use the SIPP to create a new data set of U.S. labor market behavior at weekly frequency. By comparing a measure of weekly labor force transitions with a measure of month-over-month transitions that replicates the CPS, I can identify and measure time aggregation. To assess the bias from geographic mobility, I create a new database, the Longitudinal Population Database (LPD), that utilizes all longitudinal information in the CPS. The LPD contains the complete interview history of every person surveyed by the CPS over 1976–2006 and can be used to identify persons that move separately from persons with missing observations arising from other reasons. I estimate a bound on the bias from geographic mobility by comparing the behavior of the entire CPS sample with the subset known not to have moved.

Chapter 2 uses the weekly SIPP data set to estimate the role of time aggregation in measuring gross labor force flows and unemployment dynamics. Time aggregation is substantial: gross flows estimated from monthly data understate the true number of transitions by 15–24 percent. Time aggregation in both separations to unemployment and accessions from unemployment comoves positively with the business cycle. The effect from time aggregation on separations is roughly offset by its effect on accessions, however, creating no meaningful cyclical bias in measured gross flows or hazard rates. Contrary to claims by Hall (2006) and Shimer (2007), separation hazard rates calculated from the SIPP and the CPS are strongly countercyclical and remain so after adjusting for time aggregation. In addition, the separation hazard rate contributes fully one-half of the cyclical variance of the steady-state unemployment rate after adjusting for time aggregation.

Chapter 3 assesses the implications of geographic mobility for the measurement of U.S. labor market dynamics using the CPS. It first describes constructing the LPD and introduces the longitudinal concepts necessary to identify geographic mobility. The LPD is then used to identify individuals who move into and out of the CPS sample. Comparing the populations of movers and stayers reveals little difference in the composition of sex, race, and education but large differences in age and marital status of movers compared with stayers. However, the cyclical bias arising from geographic mobility is small. At business cycle frequencies, the difference between the separation hazard rate calculated from the entire CPS sample and from a subset that is known not to have moved never exceeds 4 percent. There is little effect of mobility on the job finding hazard rate. Thus, geographic mobility does not significantly affect CPS labor market dynamics.

Chapter 4 examines the cyclical behavior of the labor market at weekly frequency and the role of time aggregation in direct employment-to-employment (EE) transitions. It first describes the SIPP and how the weekly data set is constructed, documenting the difficulties encountered creating the weekly series. Chapter 4 assesses the strengths and weaknesses of the SIPP data relative to the CPS and finds that the SIPP labor force stocks, gross flows, and cyclical dynamics compare favorably with those from the CPS. Abstracting from labor force participation, direct EE transitions account for one-half of all separations from employment. Similar estimates using the CPS are twice as large, however the CPS overstates EE transitions because of time aggregation. Separations to a new job are strongly procyclical while separations to unemployment are strongly countercyclical. The combination yields a nearly acyclical total separation rate. The weekly job finding rate is strongly procyclical.

Chapter 2

Understanding Unemployment Dynamics: The Role of Time Aggregation

2.1 Introduction

Understanding how unemployment changes over the business cycle is an important and controversial topic in economics. Although on the surface unemployment dynamics appear straightforward—unemployment rises when workers lose jobs and falls when unemployed persons find jobs—the behavior of its components over the business cycle and their relative importance continues to be an active and contentious area of research.

Beginning with Darby et al. (1986), and later Blanchard and Diamond (1990) and Davis and Haltiwanger (1990), researchers identified higher separation rates—inflows to unemployment—as the primary determinant of higher unemployment during recessions. Recently, Hall (2005, 2006) and Shimer (2005) have challenged this conventional wisdom, arguing that separations are not important for the cyclical dynamics of unemployment.[1] They claim that the separation rate is unaffected by the business cycle and that declines in job finding alone lead to increased unem-

1. Shimer released an updated version of his 2005 paper, hereafter Shimer (2007).

ployment during recessions.

There is considerable evidence that separations move countercyclically: the number of separations and the probability that an employed worker loses his job both increase during a recession.[2] To reconcile the constant-separation view with this evidence, Shimer (2007) argues that the countercyclicality in measured gross flows and hazard rates arises from time aggregation.

Problems of time aggregation arise when attempting to estimate a continuous-time relationship, such as unemployment duration, using data only available at discrete intervals.[3] In the context of gross flows, time aggregation arises because transitions among labor force states, for example from employment to unemployment, are measured by the change in a person's labor force status from one month to the next. If transitions occur at frequencies higher than a month, the monthly measurement may combine multiple transitions into a single "aggregate" transition. In particular, researchers worry about missing short spells of employment or unemployment. If these unmeasured spells occur with different propensity during a recession, then the measured gross flows will have a cyclical bias.

Shimer (2007) argues that the observed countercyclical movement of separations is the result of a bias from time aggregation: a decrease in the job finding probability during a recession indirectly raises the measured transition rate from employment to unemployment because workers who lose their jobs are more likely to experience measured spells of unemployment. Put differently, measured separations during a boom are too low because workers who lose their jobs quickly rematch with a new employer and are not recorded as unemployed. This implies a procyclical bias in separations due to time aggregation.

Shimer (2007) uses a theoretical model to relate measured monthly transition rates to underlying continuous-time hazard rates. Since this development, it has become standard practice to adjust for time aggregation.[4] Yet there is no way to observe or measure time aggregation using the standard source for labor market

2. Blanchard and Diamond (1990); Bleakley et al. (1999); Fujita and Ramey (2006, 2007, forthcoming); Fujita et al. (2007); Elsby et al. (forthcoming).

3. See Kaitz (1970); Kiefer (1988); Petersen (1991); Petersen and Koput (1992).

4. See Shimer (2005, 2007); Fujita and Ramey (2006, forthcoming); Yashiv (2007); Elsby et al. (forthcoming).

data, the Current Population Survey (CPS). Instead, corrections for time aggregation rely on the mechanical relationship between measured stocks and flows to adjust for time aggregation implicitly.[5]

I use high-frequency labor force data from the Survey of Income and Program Participation (SIPP) to identify and measure time aggregation and assess its implications for measured gross flows and hazard rates. The SIPP provides information about labor force status and job search behavior at a weekly frequency. Using this information I construct weekly measures of labor force status and week-over-week transitions.[6] Comparing the weekly transitions with a measure of month-over-month transitions that replicates the CPS, I can identify intramonth transitions that would not be observed in the CPS. This allows me to quantify time aggregation: the difference in the number of transitions in the weekly and the monthly measures.

Time aggregation in gross flows is substantial. Gross flows estimated from monthly data understate the true number of transitions by between 15 and 24 percent. Although monthly measures of gross flows capture a majority of labor market activity, roughly 20 percent of it occurs between measurement points. This speaks to the large level of time aggregation in gross flows, but not whether it contributes a cyclical bias.

The degree of time aggregation changes over time. Over the business cycle, time aggregation comoves positively with unemployment, especially for flows between employment and unemployment. This validates Shimer (2007)'s hypothesis that measured separations appear too high during recessions. However, time aggregation is equally procyclical in accessions (job finding), implying that the CPS misses accessions when the job finding is high (or, alternatively, that job finding appears too high during recessions). The cyclical effect of time aggregation on separations is thus roughly offset by its affect on accessions. Time aggregation does not impart a cyclical bias to measured gross flows. Contrary to the Hall-Shimer claim,

5. Shimer (2007) uses unemployment duration data to capture "short-term" unemployment in a 2-state model. There are no data available from the CPS that allow measuring time aggregation in a 3-state model.

6. At least since Perry (1972), many models of the labor market have taken the week as the fundamental unit of time. Indeed, the weekly frequency has become the modeling vanguard for discrete-time search and matching models. See Hagedorn and Manovskii (2008); Ramey (2008); Elsby et al. (forthcoming).

the separation hazard rate calculated from the SIPP is strongly countercyclical and remains so after adjusting for time aggregation.

Because the SIPP sample does not cover the entire period for which CPS data are available, I estimate adjustment factors for CPS gross flows using the relationship between time aggregation measured in the SIPP and the unemployment rate. This regression is then used to predict adjustment factors for the entire CPS sample period. In the CPS, separations to unemployment are strongly countercyclical. Adjusting for time aggregation reduces the cyclical correlation with unemployment by less than 10 percent. In addition, the separation hazard rate calculated from the CPS is strongly countercyclical and contributes fully one-half of the cyclical variance in the steady-state unemployment rate, both before and after adjusting for time aggregation.

Section 2.2 discusses how weekly SIPP data are used to estimate time aggregation. It also describes the method, unique to this literature, used to isolate components of the time series that move at business cycle frequency. Section 2.3 reports the estimates of time aggregation, highlighting both the average and cyclical behavior. Section 2.4 studies the effect of time aggregation on CPS gross flows and hazard rates.

2.2 Estimating Time Aggregation

This section describes the method used to estimate time aggregation. I first briefly describe the Survey of Income and Program Participation (SIPP); for a detailed description of the SIPP and additional comparisons with the Current Population Survey (CPS), see chapter 4. I then describe how information from the SIPP is used to create a weekly labor force measure. I next explain the algorithm for identifying time aggregation by comparing these two measures for the same person. The section concludes with econometric details of the aggregation and how cyclical components are isolated.

2.2.1 Survey of Income and Program Participation

The SIPP is an ongoing longitudinal survey of U.S. households. The largest organizational unit of the SIPP is the *panel*. Each panel is formed from a nationally-representative sample of individuals fifteen years of age and older selected from households in the civilian noninstitutional population. Each panel is randomly divided into 4 *rotation groups*, one of which is interviewed each month. At each interview respondents are asked to provide information about the previous four months. Unlike the CPS, the SIPP follows original household members who move.

The initial SIPP survey design called for each panel to last thirty-two months and have a target sample size of 20,000 households. A new panel was to begin each year, with multiple panels active at the same time to improved accuracy. In 1996 the SIPP underwent a substantial redesign; the overlapping panel structure was eliminated in favor of a substantially larger sample size and target panel length was increased to forty-eight months.

I use data from 12 SIPP panels: 1984, 1985, 1986, 1987, 1988, 1990, 1991, 1992, 1993, 1996, 2001, and 2004.[7] The time coverage of the SIPP panels begins in June 1983 and ends in December 2006, however there is an eight-month gap from March to October 2000, during which the SIPP did not conduct interviews for budgetary reasons. Together, the 12 SIPP panels yield longitudinal data on over 610,000 persons and cover more than twenty years. Table 2.1 presents basic statistics on the SIPP data.

In the CPS, individuals are interviewed 8 times over sixteen months, with an eight month break in between the fourth and fifth interview. Because of this break only 2 sets of 3 month-over-month labor force transitions can be measured for any person.[8] In contrast, the SIPP longitudinal data are continuous over an individual's duration in the panel, leaving only 1 unmeasurable transition. The average longitudinal duration in the SIPP sample is twenty-five months. Thus, although the SIPP has fewer people than the CPS, it contains substantially more longitudinal informa-

7. The 1989 panel contains only 3 interview waves and is not used.

8. The discrepancy between measured stocks and flows arising from unmeasurable transitions is known as "margin error" in the CPS matching literature. See Abowd and Zellner (1985); Poterba and Summers (1984, 1986); Chua and Fuller (1987); Fujita and Ramey (2006).

tion about its respondents. In addition, because the SIPP follows movers it does not suffer from geographic mobility bias.[9]

2.2.2 Synthetic CPS Labor Force Measures

I use weekly information from the SIPP to construct 2 measures of labor market transitions that, together, allow me to estimate time aggregation. One replicates how an individual from the SIPP would be classified if she was surveyed by the CPS.[10] The other adapts the CPS labor force definitions to the weekly frequency and records all weekly transitions. By comparing these two measures for the same person, I can identify and measure intramonth transitions that are not captured by the CPS.

The CPS determines an individual's labor force status for a month based on his experience during that month's *reference week*, the week of the month containing the twelfth.[11] The SIPP asks respondents to identify whether they were employed, on layoff, or searching for work in each week of the reference period. I use this information to construct a weekly measure of labor force status using CPS definitions.

Classification as employed (`E`) follows directly from the CPS definitions; unemployment and not in the labor force (NILF) are not as straightforward. The CPS classifies a person as unemployed if he has searched for a job within the last four weeks. I apply this definition on a rolling basis to determine a person's weekly labor force status. That is, a person without a job would be considered unemployed (`U`) this week if he had searched for work during any of the previous four weeks, even if he did not search this week. After four weeks without search have elapsed, a person is classified as NILF (`N`).

Labor force transitions are measured by comparing a person's labor force status in two successive time periods. I define a transition from state `i` in period $t-1$

---

9. Moscarini and Thomsson (2008) discuss geographic mobility as a source of bias in CPS gross flows.

10. See chapter 4 for details.

11. The week containing the fifth is used as the reference week for December, provided that it falls entirely within the month; otherwise the week of the twelfth is used.

to state `j` in period t as an `ij` transition observed at t. Transitions are identified at two different frequencies. A person's *weekly* labor force transition is the change in labor force status from one week to the next week. A person's *monthly* labor force transition is the change in labor force status from one CPS reference week to the next CPS reference week. This "synthetic" CPS labor force transition records how a person would have been classified by the CPS.

Although the SIPP is designed for different purposes than the CPS, the labor force statistics calculated from the SIPP match those from the CPS remarkably well.[12] Some difference between the two data sources is expected due to sampling variation and minor differences in survey design and definitions. For the most part, the SIPP and the CPS capture the same dynamics of the U.S. labor market.

The estimated population in the SIPP and in the CPS are not statistically different and the time-series correlation between the two population levels is 0.9967. The correlation of the stock of employed persons measured from the two data sources is high (0.91) as is the correlation for unemployed persons (0.94). Gross flows calculated from the CPS and the SIPP also behave similarly: the correlation for `EU` transitions is 0.83 and for `UE` transitions is 0.73. Although the SIPP and CPS agree less on measures involving persons NILF, particularly for `UN` flows, the dynamics involving unemployment are closely related in the two data sources.

2.2.3 Identifying Time Aggregation

Before describing how the parallel labor force transition series are used to identify and estimate time aggregation, it is instructive to see an example of the two measures. Table 2.2 shows the labor force history for a SIPP respondent. The first two columns show the month and week within the month. The next column shows the weekly labor force status, according to CPS definitions, for each week. The final two columns report the monthly and weekly labor force transitions. Shaded rows indicate CPS reference weeks; information would only be available for these weeks if this person was surveyed by the CPS.

This individual begins March 1990 with a job after being employed in Febru-

12. Chapter 4 compares in detail the stocks and gross flows derived from the two data sources.

ary (not shown). He remains employed through the month of March; the monthly and weekly measures are identical. In the second week of April he becomes unemployed and because that week is the CPS reference week, the CPS would classify him unemployed in April. Both the weekly and monthly measures record this transition as `eu`.

The period between the April and May reference weeks illustrates a situation where the measures differ. The monthly measure records a `ue` transition, reflecting his change from the previous reference week. The weekly transition measure also records a `ue` transition. However it also records 2 additional transitions not captured by the monthly measure. Because the CPS does not have information about the period between interviews, the `ue`, `eu`, and `ue` transitions would be "aggregated" into a single `ue` transition.

The next five weeks provides a dramatic illustration of time aggregation. Because this individual was employed in the reference weeks for May and June, the monthly measure records no transition (`ee`). Although the monthly measure records a nontransition, in fact 4 unique transitions occurred. The CPS would miss the short spells of employment, unemployment, and nonparticipation and their associated transitions.

The transitions in July and August illustrate situations where both the monthly and weekly measures record the same transition, but the week in which the transition is identified differs. Since both measures agree at some point over the period between CPS reference weeks, these cases are *not* measured as time aggregation.

Identification Algorithm

The principal objective is to measure intramonth transitions that are missed by traditional month-over-month measures. This is achieved by searching the period in between reference weeks to identify missed transitions.

The algorithm uses two sets of counters for each person. There are 9 counters $C(\mathtt{ij})$ in each set, one for each possible transition from state $\mathtt{i} \in \{\mathtt{e}, \mathtt{n}, \mathtt{u}\}$ to state $\mathtt{j} \in \{\mathtt{e}, \mathtt{n}, \mathtt{u}\}$. The first set records the traditional month-over-month labor force transition—that is, the change in labor force status from the previous CPS reference week to the current one. Only 1 counter in this first set takes on a value of 1 in a month.

The second set of counters, $C(\mathtt{ij})^*$, records all weekly transitions, including those that occur between CPS reference weeks. Transitions starting from the current month's reference week up to, but not including, the previous reference week are considered a part of the *current* month. In the example, the 5 transitions between week 3 of May 1990 and week 2 of June 1990 are counted in June. When the same type of transition occurs more than once in the period, each unique instance is counted; ee, nn, and uu transitions must be interrupted by another transition to be counted more than once. In the example from table 2.2, in May 1990 the starred counters would be: $C(\mathtt{uu})^* = 1$, $C(\mathtt{ue})^* = 2$, and $C(\mathtt{eu})^* = 1$.

Upon completion, each person has two parallel measures of her transitions for each month. The first is the traditional month-over-month labor force transition (synthetic CPS). The second records all transitions that occur between the CPS reference weeks. Differences between the two measures for the same person identify time aggregation at the individual level. However, because there is no sensible metric for time aggregation at the individual level—it is generally either zero or infinite—I study time aggregation using aggregated data.

2.2.4 Aggregation and Estimation

The aggregation of all individual ij transitions is called the IJ *flow*, where capital letters indicate the aggregate quantity. Thus IJ is the number of persons who move from state I in month $t-1$ to state J in month t as measured by the CPS. Similarly, $\mathtt{IJ}^*$ is the number of persons who make the same transition, accounting for all weekly transitions over the month. Time aggregation is defined as the ratio:

$$T_t^{\mathtt{IJ}} = \frac{\mathtt{IJ}_t^*}{\mathtt{IJ}_t}. \tag{2.1}$$

The ratio $T^{\mathtt{IJ}}$ gives the relative increase in the IJ flow resulting from measuring all intramonth transitions—that is, from time aggregation. If there were no intramonth transitions, then both transition measures would be the same and $T^{\mathtt{IJ}} = 1$. However if $\mathtt{IJ}^*$ identifies transitions not captured by IJ then $T^{\mathtt{IJ}} > 1$, indicating positive time aggregation. It is also possible to have $T^{\mathtt{IJ}} < 1$; this could occur, for example, if transitions are misclassified by time aggregation.

This ratio is an appealing metric for several reasons. First, it is scale free, allowing the bias in flows with different magnitudes to be uniformly compared. Second, the ratio construction eliminates the "seam effect" and any panel- or rotation group–specific measurement error that causes problems when comparing aggregate estimates.[13]

When estimating a longitudinal object such as gross flows, each rotation group should be thought of as its own separate panel—where here "panel" has its traditional econometric meaning: a collection of repeated observations on the same cross-section of individuals. Because each SIPP panel is nationally representative and because households are randomly assigned to rotation groups, the SIPP data can be viewed as 48 smaller, overlapping panels.

Let $p = 1, 2, \ldots, 12$ index SIPP panels and $r \in \{1, 2, 3, 4\}$ index the rotation group within a SIPP panel. An individual rotation group is uniquely identified by pr. In month t there are observations from P_t panels, each with R_{pt} rotation groups. Let $j = 1, 2, \ldots, m_{prt}$ index persons from rotation group pr in month t.

Time aggregation is estimated using a ratio estimator for population totals. The month t estimator for time aggregation in the `IJ` flow is

$$\widehat{T}_t^{\mathtt{IJ}} = \frac{\widehat{\mathtt{IJ}}_t^*}{\widehat{\mathtt{IJ}}_t} = \frac{\sum_{p=1}^{P_t} \sum_{r=1}^{R_{pt}} \sum_{j=1}^{m_{prt}} \omega_{prt} w_{prjt} \, C(\mathtt{ij})_{prjt}^*}{\sum_{p=1}^{P_t} \sum_{r=1}^{R_{pt}} \sum_{j=1}^{m_{prt}} \omega_{prt} w_{prjt} \, C(\mathtt{ij})_{prjt}}, \tag{2.2}$$

where $C(\mathtt{ij})$ and $C(\mathtt{ij})^*$ are the transition counters for person prj (discussed in section 2.2.3). Each individual's observations are weighted by their monthly sampling weight w_{prjt}. Each rotation group is weighted by its contribution to the total number of observations in a month:

$$\omega_{prt} = \frac{N_{prt}}{\sum_{p=1}^{P_t} \sum_{r=1}^{R_{pt}} N_{prt}}. \tag{2.3}$$

The pooled estimates are found by further aggregating equation 2.2 over

13. See chapter 4 for a discussion of these problems.

time:

$$\widehat{T}^{\mathtt{IJ}} = \frac{\sum_{t=1}^{T} \widehat{\mathtt{IJ}}_t^*}{\sum_{t=1}^{T} \widehat{\mathtt{IJ}}_t}. \tag{2.4}$$

Equations 2.2 and 2.4 are estimated separately for each IJ transition. The variance of the pooled-sample estimate is calculated to reflect survey sampling uncertainty; see appendix A for details.

2.2.5 Identifying Cyclical Components

Analyzing the cyclical behavior of a time series requires isolating the elements of the time series with periodic variation at a certain frequency. The traditional technique in macroeconomics is to filter the series to extract the cyclical component: the lowpass Hodrick-Prescott (HP) filter and bandpass Baxter-King (BK) filter are common methods.[14]

A drawback of such ad hoc filters is that they can lead to spurious cycles and other distortions.[15] Instead, I employ the structural time series modeling approach developed by Harvey (1989).[16] This approach views the time series as the sum of distinct unobserved components, each with an economic interpretation.

Such models are very flexible and can optimally replicate the gain properties of the HP and BK filters; in recent work, Harvey and Trimbur (2003) derive optimal lowpass and bandpass filters as the joint solution to a signal extraction problem in an unobserved-components model. An additional benefit of this approach is that missing values can be estimated directly from the structural model.

I model the times series behavior of time aggregation, and later of gross flows and hazard rates, using a structural time series model. I model the observed time series as the sum of four independent, unobserved components: a trend, a cycle, a seasonal, and an irregular component. The trend represents low-frequency movements that, when extrapolated, give the clearest indication of the future long-term movements in the series.[17] The cyclical component is a periodic function of time

14. Hodrick and Prescott (1997); Baxter and King (1999).
15. Harvey and Jaeger (1993); Cogley and Nason (1995); Murray (2003).
16. See also Durbin and Koopman (2001).
17. Harvey (1989), p.284

with a frequency at that of the business cycle. The seasonal component represents fluctuations that repeat annually and the irregular component captures the remaining non-systematic variation.

The structural time series model for the natural logarithm of each series, denoted y_t, is

$$y_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t \,, \tag{2.5}$$

where μ_t is the trend, ψ_t the cyclical, γ_t the seasonal, and ε_t the irregular component. Details of the econometric specification of the components are provided in appendix B.

Equation 2.5 is recast as a state space model where the unobserved components are represented by the state of the system. The unknown parameters are estimated by maximum likelihood using the Kalman filter to update and smooth the unobserved state. The estimation is performed using the STAMP program written by Koopman et al. (2007). The state space form and the details of the estimation appear in appendix B.

2.3 Results

I first present estimates from a pooled sample of all months. These results establish the quantitative importance of time aggregation. I then examine the cyclicality of time aggregation and explore its implications for the cyclicality of hazard rates in the SIPP.

2.3.1 Pooled-Sample Estimates of Time Aggregation

The long-term level of time aggregation in each `IJ` flow is estimated with equation 2.4 using a pooled sample of all 273 months. The pooled sample contains 15.9 million observations covering July 1983–December 2006.

Table 2.3 reports summary statistics from the pooled estimation, grouped by type of flow. One can interpret the values in table 2.3 as the percentage increase in the `IJ` flow resulting from measuring all weekly transitions. The central result is that many transitions are missed because of time aggregation. Among transitions between different labor force states, the smallest increase is 14.5 percent while the largest is 24.4 percent. These values are very precisely estimated, with standard errors of 0.36 percent or less. Thus failing to account for time aggregation substantially understates the true magnitude of gross flows.

Focusing on flows between employment and unemployment, I estimate that time aggregation in separations (23 percent) is about equal to that in accessions (24 percent). Although the large degree of time aggregation in separations is consistent with Shimer (2007), the equally large degree of time aggregation in accessions sharply conflicts; Shimer claims that time aggregation causes little bias in the job finding rate.[18] This claim is not supported by the data.

It is instructive to examine how intramonth transitions—which would not be observed in the CPS— are classified by the synthetic CPS measure. This exercise considers only cases where the weekly measure identifies a transition that the monthly measure does not. Figure 2.1 shows the distribution of the monthly classifications for each type of unrecorded weekly transition. The panel title identifies the unrecorded transition; the 9 possible monthly classifications are listed along the abscissa and their relative frequencies are plotted vertically.

The UE panel shows that unrecorded separations would be classified primarily as EE transitions in the CPS. Close to 60 percent of intramonth separations are classified as continuous employment.[19] Observations at monthly frequency miss primarily short spells of nonemployment. However, over a quarter of EU separations are incorrectly classified as UU, indicating that the CPS misses short spells of *employment* as well. The remaining nondiagonal transitions represent a combination of misclassification and multiple intramonth transitions; collectively they represent about 20 percent of transitions.

The pattern for unrecorded UE transitions, also shown in figure 2.1, is similar to that for separations to unemployment. Although about 20 percent are classified

18. Shimer (2007) p. 6.

19. I do not identify transitions directly from one employer to another without an intervening spell of unemployment. Fallick and Fleischman (2004) report that such direct employment-to-employment transitions are more than twice the magnitude of flows from employment to unemployment.

as UU, compared to 26 percent for EU transitions, only 46 percent are originally classified as EE. The difference comes from the 21 percent of UE transitions classified as NE. The misclassification hypothesis is explored below.

Misclassification from Time Aggregation

Although time aggregation in separation flows to nonparticipation (EN) has the same magnitude as in those to unemployment (table 2.3), time aggregation in accessions from nonparticipation (NE)is considerably lower than from unemployment. The 0.5-point difference in time aggregation between the two accession flows results from classification errors due to time aggregation.

Such a misclassification can arise because the CPS does not inquire about job search behavior for persons classified as employed. Thus, if a person was NILF during the previous reference week but searched for and found a job by the current reference week, the CPS would incorrectly record the transition as NE.

I employ a search procedure similar to that used to identify time aggregation to test this hypothesis. For each NE transition, I search backwards up to the previous reference week for active job search. If any job search is identified, the person must have experienced NU and UE transitions that were time-aggregated to NE.

This procedure confirms that about 5 percent of NU–UE transitions are incorrectly classified as NE. That is, the CPS incorrectly identifies too few UE transitions. Time aggregation in UE flows is 4.5 percent higher than in NE flows, consistent with the evidence that the CPS undercounts UE transitions. In addition, timing-related classification error also explains much of the difference in measured time aggregation in participation flows. If 5 percent of NE flows are actually UE flows, the CPS also misses the NU transition when the person began searching for a job.

The UE and NU panels of figure 2.1 corroborates the timing hypothesis. The bulk of unrecorded NU transitions are counted as NE, indicating that the CPS missed a short spell of unemployment in between nonparticipation and employment. Also, the UN panel shows that 36 percent of UN transitions were recorded as UE flows. These are missed UN–NE transitions, possibly arising from cessation of search after finding a job but before starting employment.

Short-Duration Spells

The existence of short spells is implied by the frequent and regular misclassification of labor force transitions using monthly data. The data show that most intramonth accessions and separations are incorrectly classified as a nontransition by the CPS. This is confirmed by the large proportion of unmeasured short spells recorded in the diagonal transitions, particularly spells of unemployment.

Table 2.3 also reports time aggregation in diagonal flows—“flows” between the same labor force state. These nontransitions comprise most of the labor market activity recorded, together accounting for 96 percent of all transitions. Time aggregation measured in these flows has the interpretation as an unrecorded short spell in that state. Time aggregation increases `EE` flows by about 5 and `NN` flows by roughly 8 percent. The striking result, however, is the 60 percent increase in `UU` flows due to time aggregation.

An important finding uncovered in the weekly SIPP data is that short spells of employment and unemployment occur considerably more frequently that previously thought. This finding is important for understanding time aggregation but also the dynamics of the U.S. labor market more broadly.

2.3.2 Cyclical Behavior of Time Aggregation

The previous section demonstrated the quantitative importance of time aggregation on average over 1983–2006. However important questions involve the time-series behavior of time aggregation. In particular, Shimer (2007) argues that failure to account for time aggregation imparts a countercyclical bias to measured `UE` flows. This point is central to his claim that the separation hazard rate is acyclical. Examining the cyclical behavior of time aggregation can inform about this claim.

Time-Series Behavior

Before analyzing the cyclical component, I first plot time-series behavior of time aggregation. A time series of the level of time aggregation is estimated separately for each of the 6 nondiagonal flows using equation 2.2. Figures 2.2–2.4 plot the time series of time aggregation along with the combined trend-cycle component estimated from the structural model (2.5), $\widehat{\mu}_t^{\mathtt{IJ}} + \widehat{\psi}_t^{\mathtt{IJ}}$. This is conceptually similar to plotting smoothed, seasonally-adjusted data. Shaded bars indicate recessions as dated by the National Bureau of Economic Research (NBER).

Figure 2.2 shows the degree of time aggregation in separation flows. Although there are several large outliers in the EN series, time aggregation in both separation series exhibits low time-series volatility: the ratio of the standard deviation to the mean is 0.08 for EU and 0.09 for EN. For comparison, the same volatility measure in gross flows is 0.19 for EU and 0.20 for EN. Time-series volatility in time aggregation is less than half as large as in gross flows. Time aggregation in EU flows exhibits a secular decline over the sample, while that in EN flows declines only through the mid-1990s.

Time aggregation in accession flows is shown in figure 2.3. The evolution of time aggregation for each accession flow follows its counterpart separation flow closely. The time-series variation of time aggregation in UE flows (0.08) is slightly lower than its separation counterpart and it exhibits the same downward trend. Time aggregation in NE flows is equally as volatile (0.11) as in EN flows, although the trend-cycle component is smoother. As with separations, time aggregation in accessions is about half as volatile as in gross flows.

Finally, time aggregation in participation flows is shown in figure 2.4. Unlike time aggregation in separations and accessions, there is no secular trend in the series. Both series exhibit low time-series volatility.

Cyclicality of Time Aggregation

Shimer (2007) argues that ignoring time aggregation will bias a researcher towards finding countercyclical separations to unemployment.[20] Shimer's argument implies that time aggregation in EU flows is procyclical. On the other hand, Fujita and Ramey (2006) find little evidence of cyclicality in their theory-based time aggregation adjustment. This section evaluates the cyclicality of time aggregation estimated from the SIPP data.

20. Shimer (2007), p. 3.

I use as cyclical indicators the civilian unemployment rate published by the Bureau of Labor Statistics (BLS) and the index of industrial production published by the Board of Governors of the Federal Reserve. Although I focus on the unemployment rate, I also present results using industrial production as an alternate indicator of the business cycle as a robustness check against the possibility that the unemployment rate is directly affected by time aggregation. In addition, industrial production is a measure of output, rather than employment, and thus its cyclical dynamics are not directly related to the *measurement* of labor market activity. That said, the correlation between cyclical components of industrial production and of the unemployment rate is very strong (-0.91). The cyclical components of the cyclical indicators are estimated from the structural model described in section 2.2.5.

Table 2.4 reports the contemporaneous correlation between the cyclical component of time aggregation and the cyclical component of each cyclical indicator. Significance of the correlation is calculated using Fisher's transformation.[21]

Focusing on EU and UE flows, the cyclical correlation of time aggregation with unemployment is negative, indicating that time aggregation in these flows is procyclical. The cyclical correlation of EU separations is -0.33, demonstrating moderate procyclicality. Although Shimer (2007) does not report cyclicality directly, this result is consistent with the direction of bias implied by his claims of acyclical separations.

The procyclical pattern of time aggregation in separations can be thought of as follows. Recessions are characterized by an increase in the number of separations to unemployment; the true number of separations, $\mathtt{EU}^*$, increases. However because the likelihood of quickly finding a new job declines in a recession—increasing the likelihood of experiencing a measured spell of unemployment—the CPS records a greater share of the true number of separations. That is, EU increases by *more* than $\mathtt{EU}^*$. Therefore the ratio $T^{\mathtt{EU}} = \mathtt{EU}^*/\mathtt{EU}$ declines during a recession, even though separations are increasing.

Unlike Shimer, however, I find that time aggregation in UE accessions is

21. Pearson's product-moment correlation coefficient, r, is not normally distributed but can be transformed by $z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$ to be asymptotically normal; see Fisher (1915).

strongly procyclical (-0.42), in fact more procyclical than in separations. Procyclicality in accessions is a natural consequence of greater time spent in unemployment. During a boom, workers who lose jobs and find new jobs quickly experience EU and UE transitions which are aggregated into an EE transition. As job finding slows during a recession and there are more measured spells of unemployment, those EU and UE transitions are now more likely to be observed separately by the CPS.

Thus, on balance, one should expect time aggregation to contribute little bias to the cyclicality of gross flows or hazard rates. This offsetting effect helps explain why Fujita and Ramey (2006) found little cyclical effect from their mechanical time aggregation correction.

Richer dynamics in the cyclical correlation are revealed by a plot of the cross-correlations between the cyclical components of time aggregation and the cyclical indicators. This shows not only the contemporaneous correlation but also how time aggregation relates to the business cycle at other horizons. I calculate the cross-correlation between the cyclical indicator in month t and $j = 0, 1, \ldots, 24$ leads and lags of time aggregation, $\text{corr}\left(\widehat{\psi}_t^{cyc}, \widehat{\psi}_{t+j}^{\mathtt{IJ}}\right)$, where $\widehat{\psi}_t$ is the cyclical component estimated from the structural times series model and $cyc = \{UR, IP\}$ is the cyclical indicator. Confidence intervals for the correlation are calculated using Fisher's transformation.

The cross-correlation of time aggregation with the unemployment rate is plotted in figure 2.5. The cross-correlation for time aggregation in EU flows has a hump shape, with peak correlation of -0.39 at $j = -4$. This confirms that time aggregation in EU flows is moderately procyclical and indicates that changes in time aggregation lead changes in unemployment by four months.

Time aggregation in UE accessions also has a hump shape, but it is less pronounced than for EU flows. The peak correlation of -0.43 occurs at a lag of three months, with a dynamic response roughly symmetric about the peak. This means that an increase in unemployment today is associated with a decline in time aggregation three months later. The peak correlation is stronger for accessions than separations.

The cross-correlations of time aggregation with the index of industrial production are shown in figure 2.6. These cross-correlations exhibit very similar patterns to those with unemployment, albeit reflected about the abscissa. The relationship with industrial production is more muted than with unemployment, with lower correlations at all leads and lags. The general patterns remain evident: a peak countercyclical correlation leading the cycle and the response of accessions shifted forward in time relative to separations. The correlations of time aggregation in EU and UE flows with industrial production are similar to those with unemployment. Both exhibit a hump shape, although the peak correlations are shifted forward in time, consistent with the cyclical component of unemployment lagging the cyclical component of industrial production.

Nonparticipation

The picture of time aggregation is less clear for flows into and out of labor force nonparticipation. NE flows are weakly procyclical when unemployment is used as a cyclical indicator but no significant relationship is found when using industrial production (table 2.4). EN flows are weekly countercyclical using either indicator. The large and similar cyclical correlations for UN and NU suggest that time aggregation in these flows arises from unmeasured short spells of unemployment.

This suggestion is confirmed by moderately procyclical time aggregation in UU flows. Procyclical short spells of unemployment is consistent with the evidence of time aggregation in EU and UE flows and with the hypothesis of increasing chance of experiencing a measurable spell of unemployment during a recession.

Time aggregation in flows between employment and nonparticipation has little significant cyclical variation when using either unemployment (figure 2.5) or industrial production (figure 2.6). Time aggregation in the participation flows is procyclical and lags the cycle. The peak correlation for NU is -0.55 at $j = 10$ while the peak correlation for UN is -0.47 at $j = 12$. The cyclical pattern of time aggregation in the participation flows is essentially the same when the index of industrial production is used as a cyclical indicator.

2.3.3 Cyclical Behavior of Hazard Rates

The previous section shows that time aggregation is procyclical in separations to unemployment but also in accessions from unemployment, confounding Shimer (2007)'s claim that time aggregation leads to a cyclical bias in the separation hazard rate. Next I directly assess the impact of time aggregation on the separation and job finding hazard rates calculated from the SIPP. This analysis also addresses the Hall-Shimer claim of a constant separation hazard rate.

I calculate monthly separation and job finding hazard rates

$$\widehat{s}_t = \frac{\mathtt{EU}_t}{\mathtt{E}_{t-1}} \quad \text{and} \quad \widehat{f}_t = \frac{\mathtt{UE}_t}{\mathtt{U}_{t-1}} \tag{2.6}$$

from the unadjusted SIPP gross flows and

$$\widehat{s}^*_t = \frac{\mathtt{EU}^*_t}{\mathtt{E}_{t-1}} \quad \text{and} \quad \widehat{f}^*_t = \frac{\mathtt{UE}^*_t}{\mathtt{U}_{t-1}} \tag{2.7}$$

from the time aggregation–adjusted gross flows, where $\mathtt{E}$ and $\mathtt{U}$ are the stock of employed and unemployed persons. I estimate equation 2.5 for each of the four hazard rates.

Figure 2.7 graphs the cyclical component (in logarithms) of the separation and job finding hazard rate against the cyclical component of the unemployment rate. The solid colored lines show hazard rates adjusted for time aggregation (equation 2.7) while dashed lines show the unadjusted hazard rates (equation 2.6). The gray line is the unemployment rate.

Looking first at the unadjusted series, it is clear from figure 2.7 that the separation hazard rate is not constant over the business cycle. Indeed, it tracks the rise in unemployment closely in each of the 2 recessions in the sample. Adjusting for time aggregation (solid line) does not alter this conclusion. The separation hazard rate is 92 percent as volatile as unemployment, falling to 79 percent after adjusting for time aggregation.

The separation hazard rate is countercyclical. The contemporaneous correlation between the cyclical components of the separation rate and unemployment is 0.43, falling slightly to 0.38 after adjusting for time aggregation. The peak correlation with unemployment, however, is strongly countercyclical, 0.71, and is basically unchanged after adjusting for time aggregation. The full cross-correlation is shown in figure 2.8.

Figure 2.7 confirms that the cyclical behavior of the job finding hazard rate closely mirrors that of unemployment. Adjusting for time aggregation does not significantly affect job finding. Consistent with previous evidence, the job finding hazard rate is more volatile than the separation hazard rate. The job finding rate is 25 percent more volatile than unemployment, *rising* to 28 percent after adjusting for time aggregation. The cross-correlation in the bottom panel of figure 2.8 shows strongly procyclical comovement with the business cycle. The contemporaneous correlation is -0.82 in the unadjusted data and -0.83 in the adjusted data; peak correlations are essentially coincident.

Motivated by Fujita and Ramey (forthcoming), I next examine how much of the variance in the unemployment rate is attributable to variations in the separation hazard rate and the job finding hazard rate. Shimer (2007) approximates the steady state unemployment rate, ur^{ss} implied by his continuous-time model by

$$ur_t \approx \frac{s_t}{s_t + f_t} \equiv ur_t^{ss}, \tag{2.8}$$

where s_t is the separation hazard rate and f_t is the job finding hazard rate.

Fujita and Ramey decompose the variation in ur_t^{ss} into factors that depend on separations and accessions plus an error term. Log linearizing ur_t^{ss} about its trend $\overline{ur}_t^{ss}$ yields the linear decomposition

$$ln\left(\frac{ur_t^{ss}}{\overline{ur}_t^{ss}}\right) = \left(1 - \overline{ur}_t^{ss}\right) ln\left(\frac{s_t}{\bar{s}_t}\right) - \left(1 - \overline{ur}_t^{ss}\right) ln\left(\frac{f_t}{\bar{f}_t}\right) + \epsilon_t, \tag{2.9}$$

where an overbar indicates a variable's trend component. Fujita and Ramey express this linearization generically as $dur_t^{ss} = dur_t^{sr} + dur_t^{jfr} + dur_t^{\epsilon}$, allowing them to decompose the total variance of dur_t^{ss} as [22]

$$\text{var}\left(dur_t^{ss}\right) = \text{cov}\left(dur_t^{ss}, dur_t^{sr}\right) + \text{cov}\left(dur_t^{ss}, dur_t^{jfr}\right) + \text{cov}\left(dur_t^{ss}, \epsilon_t\right). \tag{2.10}$$

Employing the concept of beta used in finance, they define the share of total variation, both direct and indirect, attributable to the separation rate, the job finding

22. Fujita and Ramey (forthcoming), p. 7.

rate, and the residual as

$$\beta^{sr} = \frac{\text{cov}\left(dur_t^{ss}, dur_t^{sr}\right)}{\text{var}\left(dur_t^{ss}\right)} \tag{2.11}$$

$$\beta^{jfr} = \frac{\text{cov}\left(dur_t^{ss}, dur_t^{jfr}\right)}{\text{var}\left(dur_t^{ss}\right)} \tag{2.12}$$

$$\beta^{\epsilon} = \frac{\text{cov}\left(dur_t^{ss}, \epsilon_t\right)}{\text{var}\left(dur_t^{ss}\right)}. \tag{2.13}$$

I estimate these contributions using unadjusted and time aggregation–adjusted SIPP hazard rates. The trend components in equation 2.9 are estimated using the time series model (i. e., μ_t in equation 2.5). The results of the SIPP variance decomposition are reported in table 2.5.

In the unadjusted SIPP data, variations in the separation hazard rate account for 43 percent of the variance of the steady-state unemployment rate. As suggested by figure 2.7, adjusting for time aggregation reduces the contribution of the separation rate; it contributes 38 percent of the variance in unemployment. The contribution of the separation hazard rate to the cyclical variance of unemployment is large and remains so after adjusting for time aggregation.

2.3.4 Discussion

The evidence in this section provides a clear picture that time aggregation is quantitatively important. Recording all weekly transitions in separation and accession flows increases recorded monthly flows by about 20 percent on average over 1983–2006. Although the level is high, the time series of time aggregation are roughly half as volatile as gross flows. Time aggregation exhibits more substantial volatility at business cycle frequencies. It is most pronounced for transitions into and out of unemployment, where time aggregation is moderately procyclical. Time aggregation in flows between employment and labor force nonparticipation, on the other hand, are weakly and distantly involved with the business cycle.

Although these results confirm Shimer (2007)'s argument that time aggregation in EU flows is procyclical, it is not possible to directly compare our findings because he does not report statistics about his time aggregation correction. Also, his conclusions about the cyclicality of time aggregation are based on a visual inspection of a counterfactual theoretical experiment, not on actual data. In addition, my findings about time aggregation in UE flows are sharply at odds with Shimer (2007). Whereas Shimer asserts that "time aggregation causes relatively little bias in the job finding rate," I find that, not only is the magnitude of time aggregation in UE flows higher than in EU flows, but that the cyclical correlation with unemployment is greater.[23] Therefore a researcher should be at least as concerned about time aggregation when measuring job *finding* as with job loss.

Examining separation and job finding hazard rates in the SIPP fully refutes Shimer (2007)'s claim that time aggregation imparts a countercyclical bias to the separation hazard rate. The data also refute Hall (2006)'s claim that the separation rate is constant over the business cycle. These sharply contrasting conclusions follow from two results. First, time aggregation in UE accessions is also procyclical and has a stronger contemporaneous correlation with unemployment than do EU separations. Any potential bias arising from unmeasured separations is offset by unmeasured accessions. Second, the overall volatility of time aggregation is small. Even if there were large cyclical variations in time aggregation, there is little room for time aggregation to have a dramatic effect on the cyclicality of gross flows or hazard rates.

The next section assesses the cyclical behavior of time aggregation–adjusted gross flows and hazard rates estimated from the CPS. This robustness exercise has a longer sample (1976–2007) and allows for direct comparison with pervious studies of time aggregation in the CPS.[24]

2.4 Adjusting the CPS for Time Aggregation

In this section I construct time aggregation–adjusted CPS gross flows and hazard rates and study their cyclical properties. Because the SIPP sample does not cover the entire period for which CPS data are available, I first estimate the rela-

23. Shimer (2007), p. 6.
24. Shimer (2007); Fujita and Ramey (2006, forthcoming); Elsby et al. (forthcoming).

tionship between time aggregation and the unemployment rate. This relationship is then used to predict an adjustment factor for the entire CPS sample period.

2.4.1 Time Aggregation–Adjusted CPS Gross Flows

For each flow, I regress the log of $\widehat{T}_t^{IJ}$ on the trend, cycle, and seasonal components of the unemployment rate (also expressed in logarithms),[25]

$$\ln\left(\widehat{T}_t^{IJ}\right) = \theta_0 + \theta_\mu \widehat{\mu}_t^{UR} + \theta_\psi \widehat{\psi}_t^{UR} + \theta_\gamma \widehat{\gamma}_t^{UR} + \epsilon_t. \tag{2.14}$$

The time series $\widehat{T}^{IJ}$ used as the dependent variable is estimated directly from the SIPP data using equation 2.2. The cyclical components of the unemployment rate are estimated using the structural time series model in section 2.2.5. The regressions use data for July 1983–November 2006.

The results of the time aggregation adjustment factor regressions are reported in table 2.6. The coefficient on the cyclical component, θ_ψ, is negative in all regressions, consistent with the correlations in column 1 of table 2.4. I calculate time aggregation adjustment factors as the fitted values from the regression (2.14) taken over the full time series of the unemployment rate. Figure 2.9 graphs the time series of the time aggregation adjustment factor for each of the 6 flows over 1976–2007.

The final step is to apply the time aggregation adjustment to gross flows estimated from the CPS. CPS gross flows are first adjusted for margin error; see appendix C for details. For the remainder of this chapter "unadjusted" CPS flows refer to flows not adjusted for time aggregation. The margin error-adjusted gross flows are then adjusted for time aggregation using the factors shown in figure 2.9. I then estimate the structural time series model (equation 2.5) separately for unadjusted and adjusted CPS gross flows.[26]

25. A specification including the irregular component was rejected by a likelihood ratio test; there is little reason to believe that the irregular component has predictive power.

26. It is well documented that changes in the CPS identification records prohibit matching for several months in the CPS sample. See Bleakley et al. (1999); Fallick and Fleischman (2004); Nagypál (2004); Fujita and Ramey (2006, 2007); Shimer (2007). These missing observations are estimated directly from the structural model.

Figure 2.10 graphs the combined trend and cycle components of the CPS gross flows, with and without adjusting for time aggregation, reported as a percentage of the population. Although a level shift is readily apparent from looking at the data in figure 2.10, nothing of consequence can be determined about how time aggregation affects cyclicality by visual inspection.

2.4.2 Cyclical Behavior of CPS Gross Flows

To properly assess the effect of time aggregation on the cyclicality of gross flows I focus on only the cyclical components estimated from equation 2.5.

Adjusting for time aggregation reduces the time-series volatility of the cyclical component of CPS gross flows. The cyclical volatility of EU flows and UE flows decreases by 3 and 8 percent, respectively. The volatility of gross flows relative to unemployment or industrial production is largely unchanged by time aggregation. In the unadjusted data, gross flows are 54 percent as volatile as unemployment and 2.2 times as volatile as industrial production. After adjusting for time aggregation, gross flows are 50 percent as volatile as unemployment and 2 times more volatile than industrial production.

In unadjusted gross flows separations to unemployment are 38 percent more volatile than accessions from unemployment, qualitatively consistent with Fujita and Ramey (2006)'s findings. After adjusting for time aggregation EU flows are 47 percent more volatile than UE flows. Adjusting for time aggregation *increases* cyclical volatility of separations relative to accessions—exactly the opposite conclusion from Shimer (2007).

Table 2.7 reports the contemporaneous correlation of the cyclical component of gross flows with the cyclical components of the unemployment rate and the index of industrial production. Looking first at the relationship between unadjusted gross flows and the unemployment rate, two things are apparent. First, the relationships are very strong; the weakest correlation is 0.82. Second, EU separation flows are strongly countercyclical; its contemporaneous correlation with unemployment is 0.84. The correlation for UE accessions is even stronger (0.91).

Gross flows adjusted for time aggregation yield the same conclusions as with the unadjusted data. The cyclical correlation of the time aggregation–adjusted EU separation flow with unemployment falls, from 0.84 to 0.77, but still indicates strongly countercyclical comovement. Adjusting for time aggregation lowers the correlation of UE accession flows with unemployment from 0.91 to 0.82. After adjusting gross flows for time aggregation, they still remain strongly countercyclical. One draws the same conclusions when industrial production is used as a cyclical indicator.

The minimal impact of time aggregation is clear in the cross-correlations of gross flows. Figures 2.11 and 2.12 show the cross-correlations between the cyclical component of CPS gross flows and the cyclical component of the two cyclical indicators. Each panel plots 2 series, one using data not adjusted for time aggregation (dashed line) and the other using time aggregation–adjusted data (solid line). This figure informs both upon the cyclical behavior of CPS gross flows (the solid line) and the contribution of time aggregation (the difference between the dashed and solid lines).

In figure 2.11 the peak correlation of unadjusted EU separations is strongly countercyclical (0.95) and leads the cycle by five months. Adjusting for time aggregation reduces the peak correlation to 0.90 at a lead of six months. Accessions from unemployment are strongly countercyclical (0.91) and lag unemployment by one month. Adjusting for time aggregation reduces the cyclicality of UE flows (0.83) at a lag of two months. Separations to and accessions from unemployment are strongly correlated with the business cycle and remain so after adjusting for time aggregation.

Consistent with findings by Blanchard and Diamond (1990) and Fujita and Ramey (2006), the comovement of flows to and from unemployment is almost exactly opposite that of movement to and from NILF over the business cycle: separations to NILF are strongly procyclical while accessions from NILF are strongly countercyclical. Unadjusted separations to NILF are procyclical and lag the cycle by two months. Adjusting for time aggregation reduces the peak correlation slightly, from -0.86 to -0.83, but does not affect the phasing. Job finding from NILF is strongly procyclical (-0.82) and is coincident with the business cycle; it remains procyclical (−0.79) after adjusting for time aggregation.

Participation flows are countercyclical and lag unemployment by one to two months. Adjusting for time aggregation lowers the peak correlation of NU flows from 0.84 to 0.79 and reduces it from 0.85 to 0.82 for UN flows. Both remain roughly coincident with unemployment after adjusting for time aggregation.

Figure 2.12 plots the cross-correlations of gross flows with the index of industrial production. The general pattern is the same as with unemployment. Separations to and accessions from unemployment are strongly countercyclical and become slightly less so after adjusting for time aggregation. Similarly, separations to and accessions from NILF are strongly procyclical; adjusting for time aggregation reduces the cyclical comovement negligibly. The phasing of the correlations with industrial production is shifted slightly forward in time relative to unemployment, consistent with unemployment lagging industrial production.

Although adjusting CPS gross flows for time aggregation reduces their contemporaneous and peak cyclical correlations, it does not meaningfully affect their degree or pattern of comovement with unemployment or industrial production. This finding is sharply at odds with Shimer (2007).

2.4.3 Cyclical Behavior of CPS Hazard Rates

This final section explores the effect of time aggregation on CPS hazard rates. It also explores the relationship between the data-based time aggregation correction and the mechanical correction employed by previous researchers.

As in section 2.3.3, I calculate monthly separation and job finding hazard rates from CPS gross flows with and without adjusting for time aggregation. In addition, I evaluate the mechanical correction suggested by Shimer (2007) that links month-over-month gross flows to underlying continuous-time adjustment equations. The separation hazard rate $\tilde{s}_t$ and the job finding hazard rate $\tilde{f}_t$ will satisfy[27]

$$\tilde{s}_t = \frac{s_t\left(1 - \mathrm{e}^{-\left(s_t + f_t\right)}\right)}{s_t + f_t} \qquad \text{and} \qquad \tilde{f}_t = \frac{f_t\left(1 - \mathrm{e}^{-\left(s_t + f_t\right)}\right)}{s_t + f_t}, \tag{2.15}$$

27. See Fujita and Ramey (forthcoming), p. 4.

where s_t and f_t are given by equation 2.6.

Figure 2.13 graphs the cyclical component (in logarithms) of the hazard rates against the cyclical component of the unemployment rate. The gray line is the unemployment rate. The short dashed line is the monthly hazard rate calculated from the CPS gross flows using equation 2.6. The long dashed line is the theoretical adjustment for time aggregation (equation 2.15). Finally, the solid colored line is the hazard rate adjusted for time aggregation using the data (equation 2.7).

As with the SIPP data, the separation hazard rate is obviously not constant over the business cycle. Adjusting for time aggregation using either method does not change this conclusion. The separation hazard rate is strongly countercyclical. The contemporaneous correlation between the cyclical components of the separation rate and unemployment is 0.88 and the peak correlation is 0.95 at a lead of four months (figure 2.14).

Also like the SIPP data, the job finding hazard rate comoves closely with unemployment and time aggregation has virtually no effect on measured cyclicality. The cross-correlation in the bottom panel of figure 2.14 shows strongly procyclical comovement coincident with the unemployment.

Adjusting for time aggregation makes little difference for the cyclicality of the separation and job finding hazard rates. The cross-correlations shown in figure 2.14 indicate that the theoretical time-aggregation adjustment does remarkably little to the cyclical relationships. This is consistent with the mechanical nature of the correction; it uses no new information. The data-based adjustment for time aggregation slightly reduces the cyclical correlation of the separation hazard rate and slightly increases the correlation of the job finding hazard rate.

I also perform the unemployment variance decomposition using the CPS data; table 2.8 reports the results. In the unadjusted data, variations in the separation hazard rate and job finding rate each account for one-half of the variance of the steady-state unemployment rate. Adjusting for time aggregation using the SIPP data makes virtually no difference in the variance decomposition. After adjusting for time aggregation, fully one-half of steady-state unemployment volatility results from separations to unemployment.

Shimer (2007)'s theoretical correction for time aggregation, however, reduces the contribution of fluctuations in the separation rate to 42 percent of the variance of unemployment. His correction spuriously reduces the contribution of the separation hazard rate by almost 20 percent. Thus, rather than improve estimates by adjusting for time aggregation, Shimer (2007)'s correction *biases* them toward finding lower separation volatility.

2.5 Conclusion

This chapter uses high-frequency data from the SIPP to estimate the degree to which measured CPS gross flows are biased due to the monthly sampling frequency of the survey. Economists worry—correctly—that a month may be too long an interval over which to measure the change in labor force states. By identifying and measuring transitions that happen in the weeks between interviews, I can empirically quantify and evaluate time aggregation.

Using the SIPP's weekly information on labor force status, I measure labor force transitions that are missed when information is available only once a month. I quantify time aggregation as the increase in gross flows resulting from measuring transitions that occur between interviews.

I find that the level of time aggregation is substantial. Gross flows estimated from monthly data understate the true number of transitions by between 15 and 24 percent. Although monthly measures of gross flows capture a majority of labor market activity, roughly 20 percent of it occurs between measurement points. Although the level is high, it has comparatively low time-series volatility.

Time aggregation varies over the business cycle, especially in transitions into and out of unemployment. Time aggregation in these flows is procyclical: as spells of unemployment become longer and more frequent during a recession, flows into and out of unemployment that are recorded by the CPS increase by *more* than the true flows do because more short spells of unemployment are captured by the CPS. Time aggregation in flows between employment and labor force nonparticipation, on the other hand, is weakly and distantly involved with the business cycle.

This chapter also makes a methodological contribution to the cyclical analysis of labor market behavior. Whereas previous research has extracted business-cycle frequencies using an ad hoc mix of seasonal adjustment and filtering, I isolate cyclical components in a unified model that jointly identifies unobserved components of a time series as the optimal solution to a signal-extraction problem.

There has been substantial debate in the literature about the cyclical pattern and relative importance of job separations and accessions over the business cycle.[28] Recently, Shimer (2005) has emphasized the importance of time aggregation for assessing cyclical patterns, arguing that failing to adjust for time aggregation causes separations to appear spuriously countercyclical.

The evidence presented in this chapter refutes this claim. Time aggregation in separations to unemployment comoves positively with the business cycle, consistent with Shimer's claims. Contrary to Shimer's claim, not only is the magnitude of time aggregation in job finding higher than in separations, but the cyclical correlation with unemployment is greater. After adjusting for time aggregation, the monthly separation hazard rate is strongly countercyclical and is 79 percent as volatile as unemployment over the business cycle.

Estimates from the SIPP are used to construct a dynamic time aggregation adjustment factor for CPS gross flows over 1976–2007. Adjusting for time aggregation generally reduces the cyclical volatility of CPS gross flows but *increases* the volatility of separations relative to accessions. Although adjusting for time aggregation does reduce the contemporaneous and peak cyclical correlation of gross flows, it does not meaningfully affect the degree or pattern of comovement with unemployment or industrial production. Separations are strongly countercyclical after adjusting for time aggregation.

Additionally, properly adjusting for time aggregation does not alter the contribution of the separation hazard rate. Adjusting for time aggregation using Shimer (2007)'s theoretical correction, however, spuriously reduces the contribution of the separation hazard rate. The time aggregation–adjusted CPS separation hazard rate is strongly countercyclical and contributes one-half of the cyclical variance in the

28. Darby et al. (1986); Hall (2006); Shimer (2005); Fujita and Ramey (2006, 2007, forthcoming); Fujita et al. (2007); Yashiv (2007); Elsby et al. (forthcoming).

steady-state unemployment rate.

Shimer (2007) argues that "ignoring time aggregation will bias a researcher towards finding a countercyclical employment exit probability."[29] This chapter refutes Shimer's claim: time aggregation imparts no meaningful cyclical bias to either gross flows or hazard rates. Nevertheless, although time aggregation is not important for cyclical dynamics, researchers must account for time aggregation in levels, such as when calibrating a weekly matching model.

29. Shimer (2007), p. 3.

Table 2.1. The Survey of Income and Program Participation

			Number of		
Panel	*Begin*	*End*	*Months*	*Persons*	*Observations*[a]
1984	Jun 1983	Apr 1986	35	48,498	1,077,059
1985	Oct 1984	Jul 1987	34	33,231	730,946
1986	Oct 1985	Mar 1988	30	27,215	588,511
1987	Oct 1986	Apr 1989	31	27,262	618,268
1988	Oct 1987	Dec 1989	27	26,895	516,829
1990	Oct 1989	Aug 1992	35	52,220	1,321,940
1991	Oct 1990	Jul 1993	35	33,438	848,159
1992	Oct 1991	Dec 1994	39	46,747	1,307,685
1993	Oct 1992	Dec 1995	39	46,659	1,296,200
1996	Dec 1995	Feb 2000	51	88,798	2,892,975
2001	Oct 2000	Dec 2003	39	79,834	1,948,077
2004	Oct 2003	Dec 2006	39	99,877	2,527,403
All	Jun 1983	Dec 2006	276	610,674	15,674,052

Source: Author’s tabulations using SIPP microdata for 1983:6–2006:12.

a. Monthly.

Table 2.2. Example Labor Force History[a]

		Labor	Transition	
Month	*Week*	*force status*	*Monthly*	*Weekly*
1990:3	1	e		ee
1990:3	2	e	ee	ee
1990:3	3	e		ee
1990:3	4	e		ee
1990:4	1	e		ee
1990:4	2	u	eu	eu
1990:4	3	u		uu
1990:4	4	e		ue
1990:5	1	u		eu
1990:5	2	e	ue	ue
1990:5	3	n		en
1990:5	4	e		ne
1990:5	5	u		eu
1990:6	1	u		uu
1990:6	2	e	ee	ue
1990:6	3	e		ee
1990:6	4	u		eu
1990:7	1	u		uu
1990:7	2	u	eu	uu
1990:7	3	u		uu
1990:7	4	u		uu
1990:8	1	e		ue
1990:8	2	e		ee
1990:8	3	e	ue	ee
1990:8	4	e		ee
1990:8	5	e		ee
1990:9	1	e		ee
1990:9	2	e	ee	ee
1990:9	3	e		ee
1990:9	4	e		ee

Source: SIPP microdata, 1990 panel.
a. Shading indicates CPS reference week.

Table 2.3. Time Aggregation, Pooled Estimates, 1983–2006 [a]

Flow	$\widehat{T}^{\mathtt{IJ}}$	*Standard error* [b]	*95 percent confidence interval*	
Separation				
`EU`	1.2298	0.0036	1.2228	1.2368
`EN`	1.2330	0.0025	1.2281	1.2379
Accession				
`UE`	1.2438	0.0031	1.2378	1.2498
`NE`	1.1902	0.0026	1.1851	1.1952
Participation				
`UN`	1.1455	0.0025	1.1407	1.1504
`NU`	1.2119	0.0026	1.2067	1.2171
Diagonal				
`EE`	1.0525	0.0002	1.0522	1.0529
`UU`	1.5991	0.0030	1.5933	1.6050
`NN`	1.0786	0.0004	1.0779	1.0794

Source: Author's calculations using SIPP microdata for 1983:7–2006:12.

a. Measure of time aggregation, $\widehat{T}^{\mathtt{IJ}} = \widehat{\mathtt{IJ}}^{*}/\widehat{\mathtt{IJ}}$, estimated using pooled sample of 15,947,129 observations over 273 months.

b. Linearized standard error estimated from survey data. See appendix A for details.

Table 2.4. Contemporaneous Correlation of Time Aggregation with Cyclical Indicators[a]

Flow	*Unemployment rate*	*Industrial production*
Separation		
EU	−0.3258***	0.2454***
EN	−0.1006**	0.0128
Accession		
UE	−0.4158***	0.2154***
NE	0.0822*	−0.1098**
Participation		
UN	−0.2563***	0.1928***
NU	−0.4152***	0.2878***
Diagonal		
EE	0.3606***	−0.2653***
UU	−0.4836***	0.4597***
NN	0.1394**	−0.0674

Source: Author's calculations using data from the SIPP, the BLS, and the Board of Governors of the Federal Reserve System.

a. Cyclical components estimated using equation 2.5. *** indicates significance at 1 percent, ** at 5 percent, and * at 10 percent.

Table 2.5. Contributions to Unemployment Fluctuations, SIPP, 1983–2002[a]

Component	*Unadjusted*	*Adjusted*[b]
Separation hazard rate (β^{sr})	0.4345	0.3820
Job finding hazard rate (β^{jfr})	0.5696	0.6207
Residual (β^{ϵ})	−0.0041	−0.0027

Source: Author's calculations using SIPP data for 1983:7–2006:11.

a. Share of variance of steady-state unemployment rate; see text for details.

b. Adjusted for time aggregation.

Table 2.6. Time Aggregation Adjustment Factor Regressions[a]

	Flow					
Coefficient	EN	EU	NE	UE	NU	UN
θ_0	0.8603***	1.0040***	0.7467***	0.8869***	0.1206	0.2022***
	(0.0774)	(0.0592)	(0.0747)	(0.0556)	(0.0778)	(0.0639)
θ_μ	0.2274***	0.2824***	0.2002***	0.2368***	−0.0257	0.0219
	(0.0270)	(0.0207)	(0.0261)	(0.0194)	(0.0272)	(0.0223)
θ_ψ	−0.1245***	−0.1637***	−0.0386	−0.1532***	−0.1237***	−0.0887**
	(0.0444)	(0.0339)	(0.0428)	(0.0319)	(0.0446)	(0.0366)
θ_γ	0.3579***	0.0554	0.3048***	0.2799***	0.1968***	0.3479***
	(0.0743)	(0.0565)	(0.0717)	(0.0534)	(0.0747)	(0.0614)
Summary statistic						
No. obs.	281	280	281	281	281	281
R^2	0.2575	0.4150	0.2176	0.3999	0.0563	0.1213

Source: Author's regressions using data from the SIPP and the BLS.

a. Regression of $\ln\left(\widehat{T}_t^{\mathtt{IJ}}\right) = \theta_0 + \theta_\mu \widehat{\mu}_t^{UR} + \theta_\psi \widehat{\psi}_t^{UR} + \theta_\gamma \widehat{\gamma}_t^{UR} + \epsilon_t$, where $\widehat{\mu}_t^{UR}$, $\widehat{\psi}_t^{UR}$, and $\widehat{\gamma}_t^{UR}$ are the estimated trend, cycle, and seasonal components of the unemployment rate (see section 2.2.5). Regressions use data from 1983:7–2006:11. Standard errors are reported in parentheses. * indicates significance at 10 percent, ** at 5 percent, and *** at 1 percent.

Table 2.7. Contemporaneous Correlation of Gross Flows with Cyclical Indicators, 1976–2007[a]

	Unemployment rate		*Industrial production*	
Flow	*Unadjusted*	*Adjusted*[b]	*Unadjusted*	*Adjusted*[b]
Separation				
EU	0.8426	0.7694	−0.8630	−0.8175
EN	−0.8404	−0.8173	0.7316	0.6893
Accession				
UE	0.9055	0.8147	−0.8574	−0.7946
NE	−0.8237	−0.7881	0.7747	0.7227
Participation				
UN	0.8201	0.7874	−0.7662	−0.7394
NU	0.8320	0.7783	−0.7776	−0.7314

Source: Author's calculations using data from the SIPP, the BLS, and the Board of Governors of the Federal Reserve System.

a. Cyclical components estimated using equation 2.5.

b. Adjusted for time aggregation using the adjustment factors in figure 2.9.

Table 2.8. Contributions to Unemployment Fluctuations, CPS, 1976–2007[a]

		Adjusted	
Component	*Unadjusted*	*Data*[b]	*Theoretical*[c]
Separation hazard rate (β^{sr})	0.4982	0.5021	0.4224
Job finding hazard rate (β^{jfr})	0.5090	0.5046	0.5850
Residual (β^{ϵ})	−0.0073	−0.0068	−0.0073

Source: Author's calculations using data from the SIPP and the CPS.

a. Share of variance of steady-state unemployment rate; see text for details.

b. Adjusted for time aggregation using the adjustment factors in figure 2.9.

c. Adjusted for time aggregation using Shimer (2007)'s theoretical model.

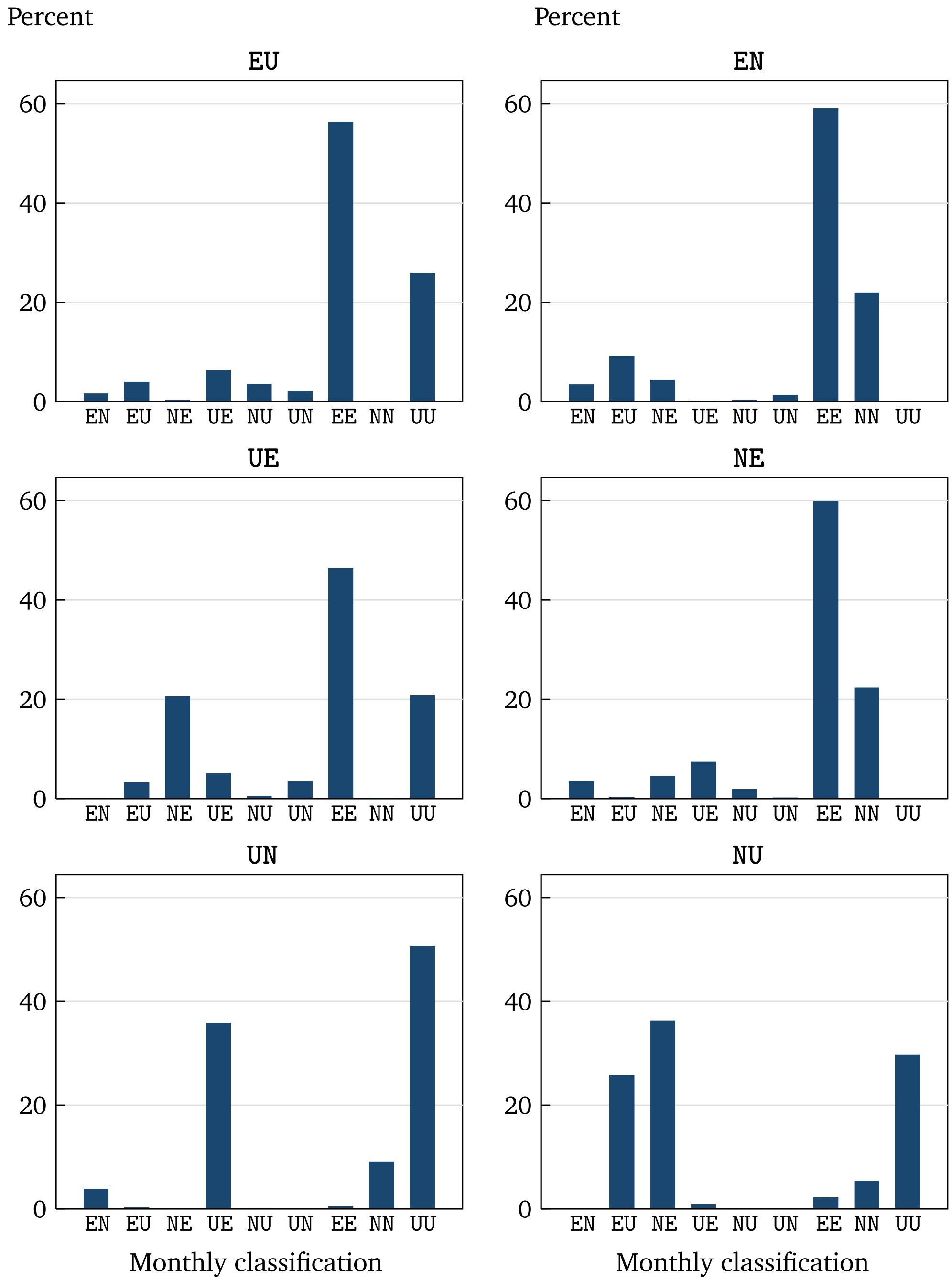

Figure 2.1. Origins of Time Aggregation[a]

Source: Author's calculations using weekly SIPP data for 1983:6–2006:12.
a. Distribution of monthly classification for unrecorded weekly transitions.

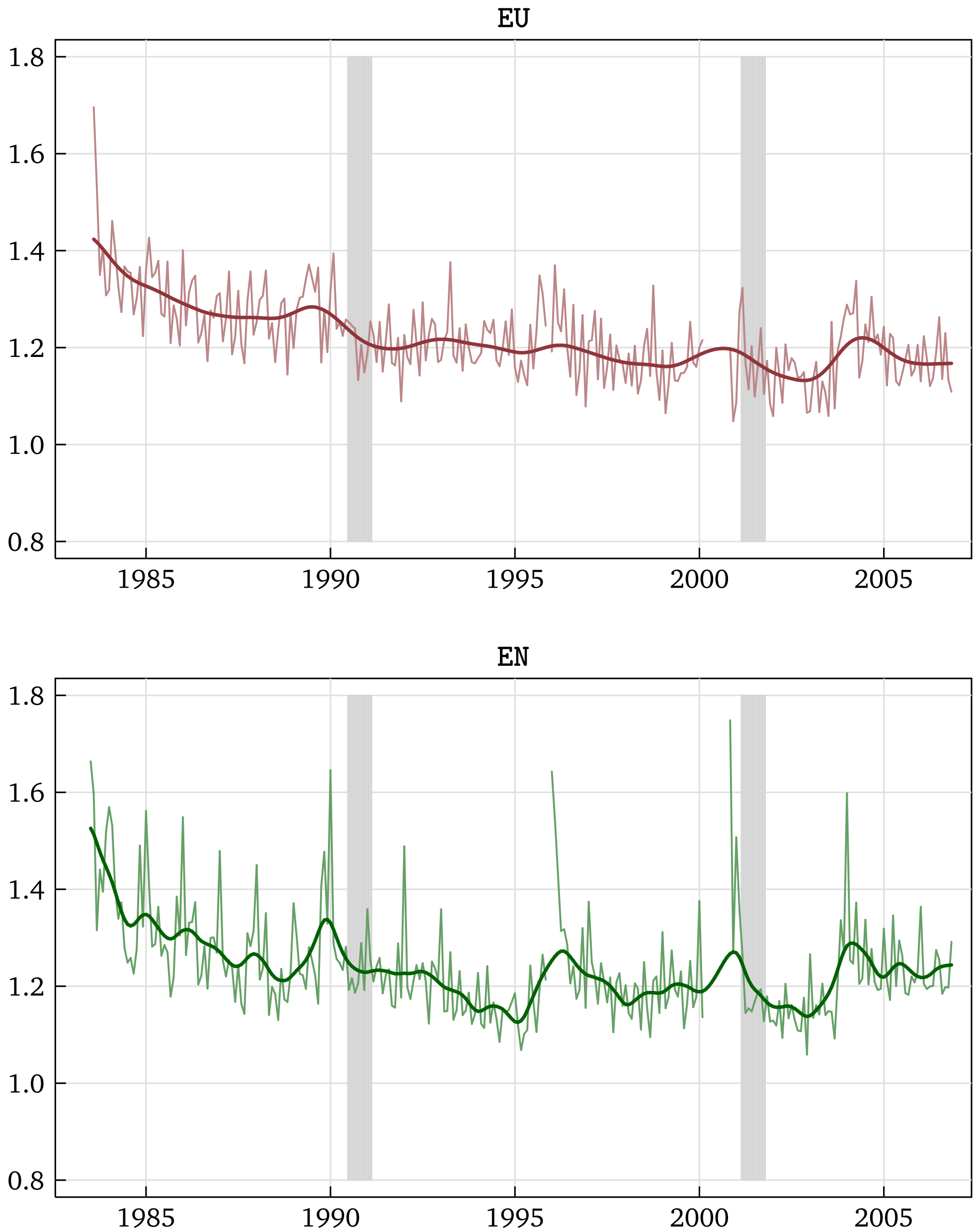

Figure 2.2. Time Aggregation, Separation Flows, 1983–2006[a]

Source: Author's calculations using SIPP microdata for 1983:7–2006:11.

a. Graph of $\widehat{T}_t^{\mathtt{IJ}} = \widehat{\mathtt{IJ}}_t^* / \widehat{\mathtt{IJ}}_t$ shown with combined trend-cycle component estimated using equation 2.5. Shaded bars indicate NBER-dated recessions.

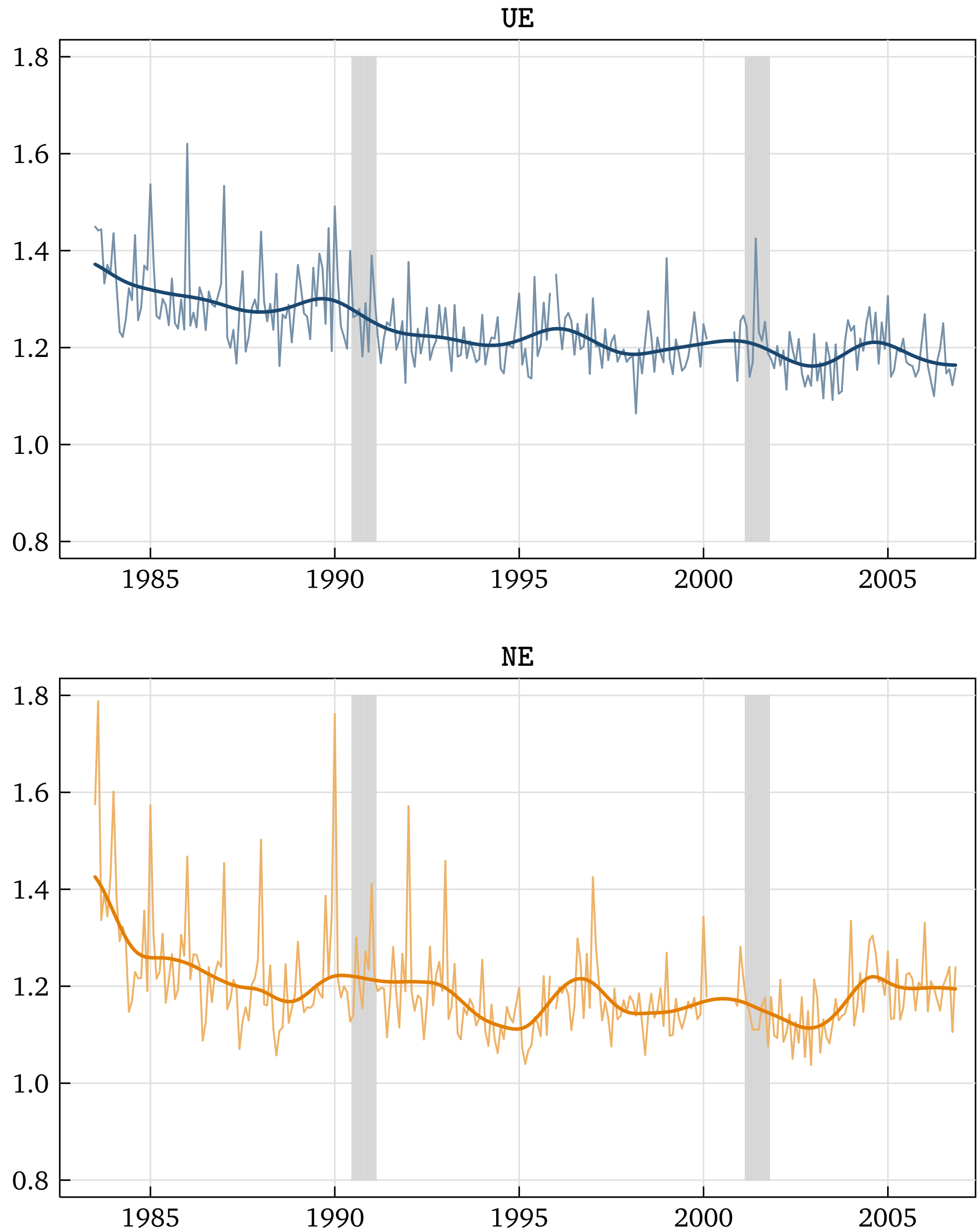

Figure 2.3. Time Aggregation, Accession Flows, 1983–2006[a]

Source: Author's calculations using SIPP microdata for 1983:7–2006:11.

a. Graph of $\widehat{T}_t^{\mathtt{IJ}} = \widehat{\mathtt{IJ}}^*_t / \widehat{\mathtt{IJ}}_t$ shown with combined trend-cycle component estimated using equation 2.5. Shaded bars indicate NBER-dated recessions.

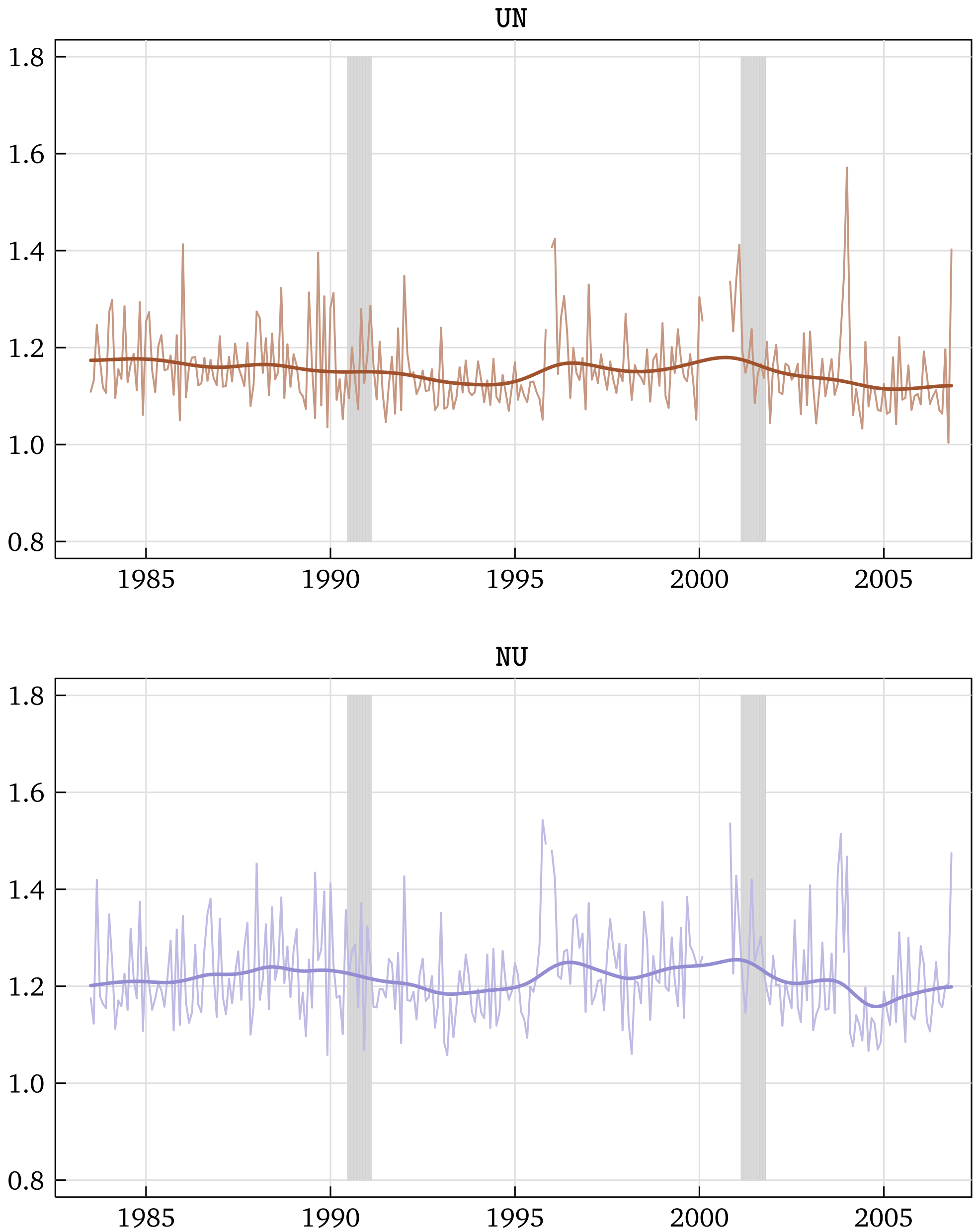

Figure 2.4. Time Aggregation, Participation Flows, 1983–2006[a]

Source: Author's calculations using SIPP microdata for 1983:7–2006:11.

a. Graph of $\widehat{T}_t^{\mathtt{IJ}} = \widehat{\mathtt{IJ}}_t^{*}/\widehat{\mathtt{IJ}}_t$ shown with combined trend-cycle component estimated using equation 2.5. Shaded bars indicate NBER-dated recessions.

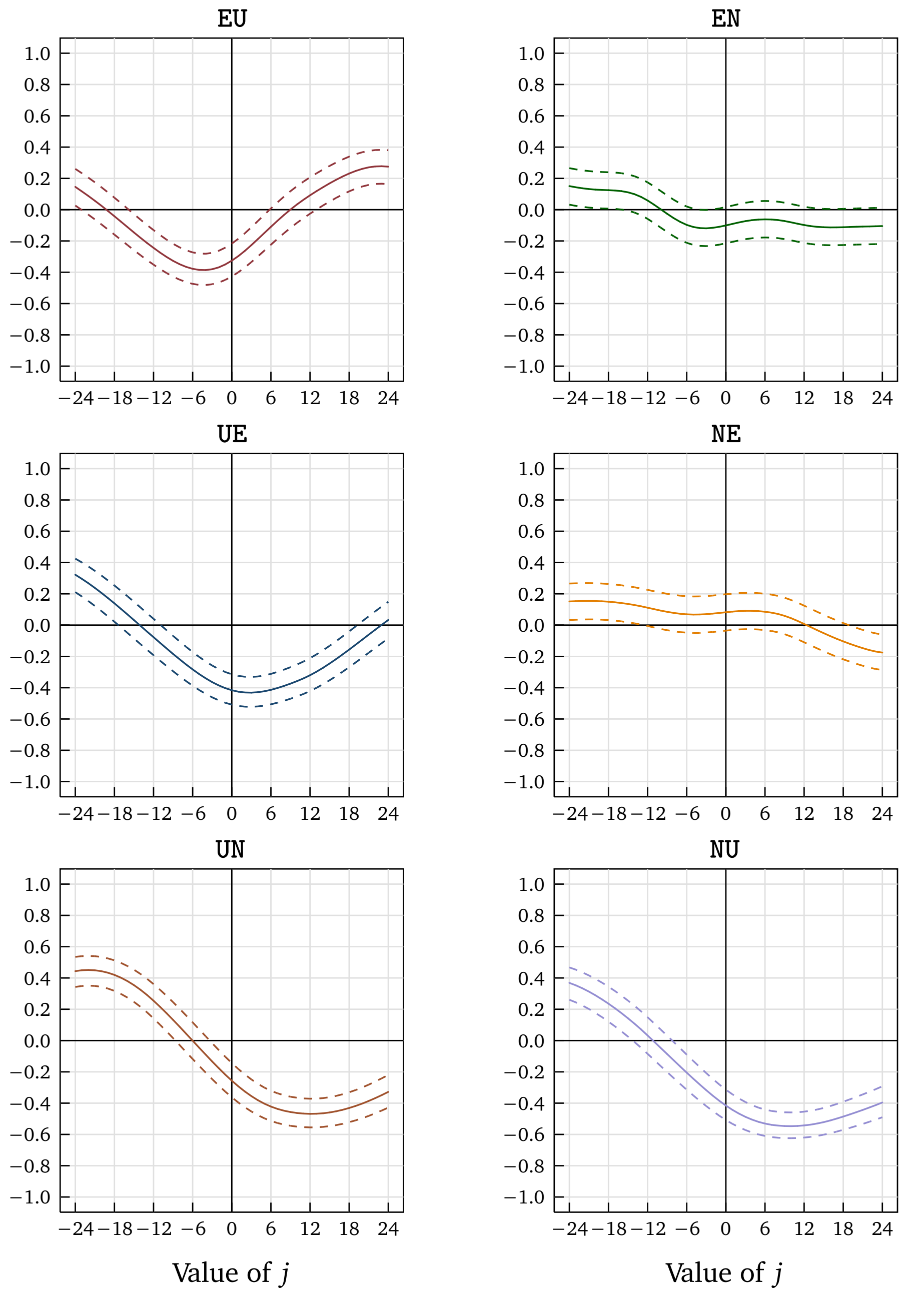

Figure 2.5. Cross-Correlations of Time Aggregation with Unemployment Rate, 1983–2006[a]

Source: Author's calculations using data from the SIPP and the BLS.

a. Correlation of $\widehat{\psi}_t^{UR}$ with $\widehat{\psi}_{t+j}^{T^{IJ}}$. Cyclical component estimated using equation 2.5. Dashed lines indicate 95 percent confidence interval.

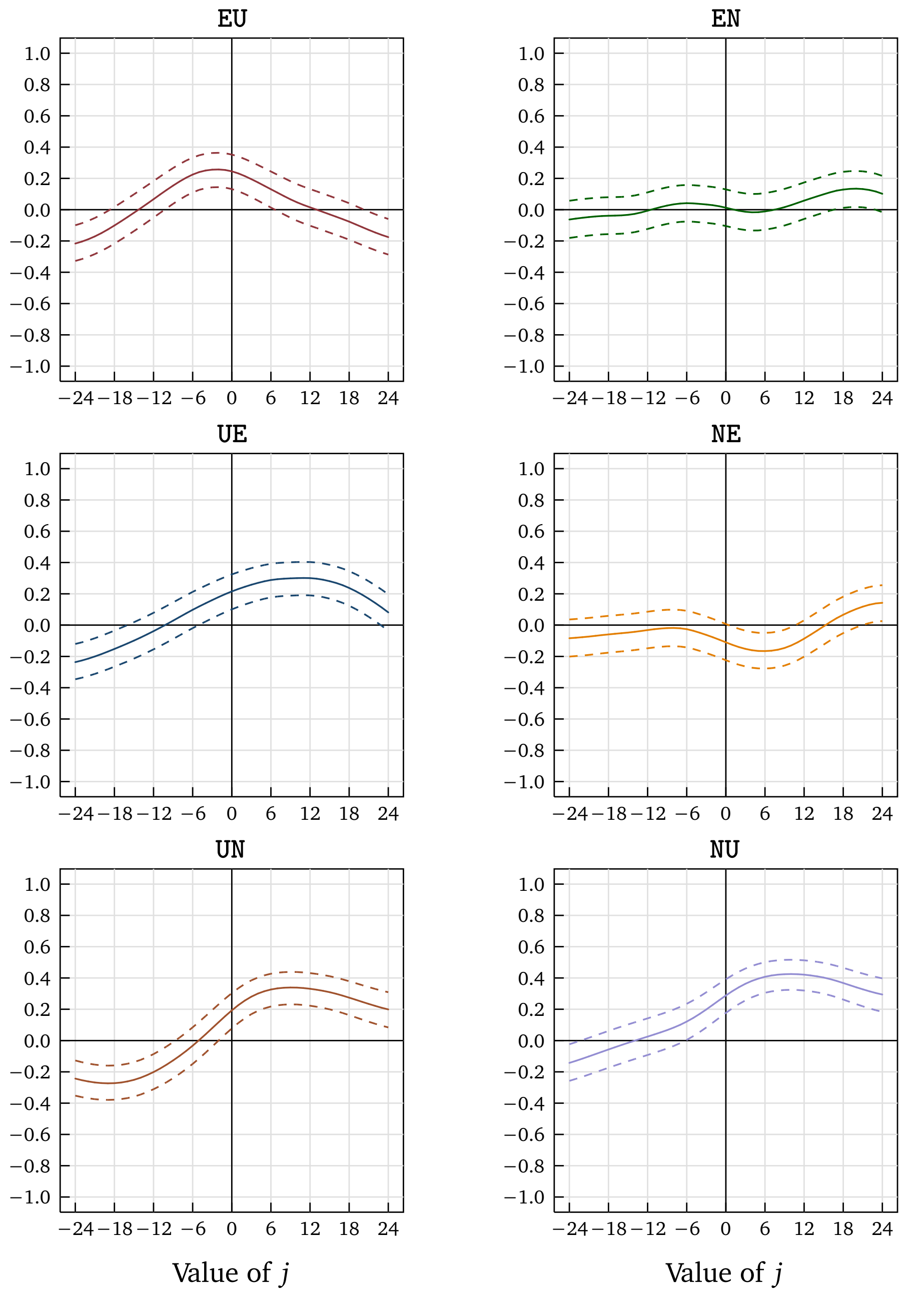

Figure 2.6. Cross-Correlations of Time Aggregation with Industrial Production, 1983–2006[a]

Source: Author's calculations using data from the SIPP and the Board of Governors of the Federal Reserve.

a. Correlation of $\widehat{\psi}_t^{IP}$ with $\widehat{\psi}_{t+j}^{T^{IJ}}$. Cyclical component estimated using equation 2.5. Dashed lines indicate 95 percent confidence interval.

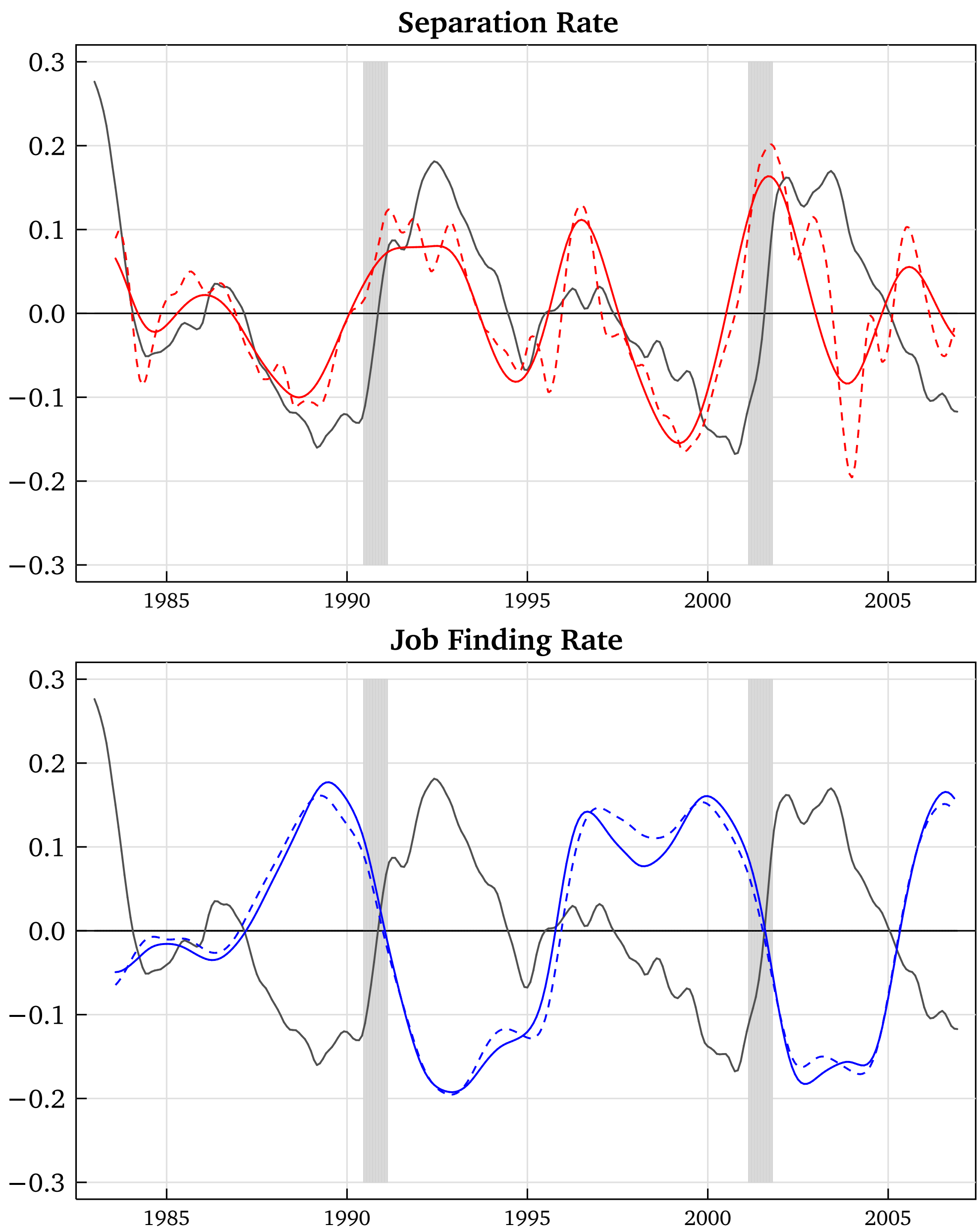

Figure 2.7. Cyclical Component of Separation and Job Finding Hazard Rates, SIPP, 1983–2006[a]

Source: Author's calculations using SIPP microdata for 1983:7–2006:11.

a. Cyclical component estimated using equation 2.5. Dark gray line is cyclical component of unemployment rate. Solid colored lines are adjusted for time aggregation; dashed lines are unadjusted. Shaded bars indicate NBER-dated recessions.

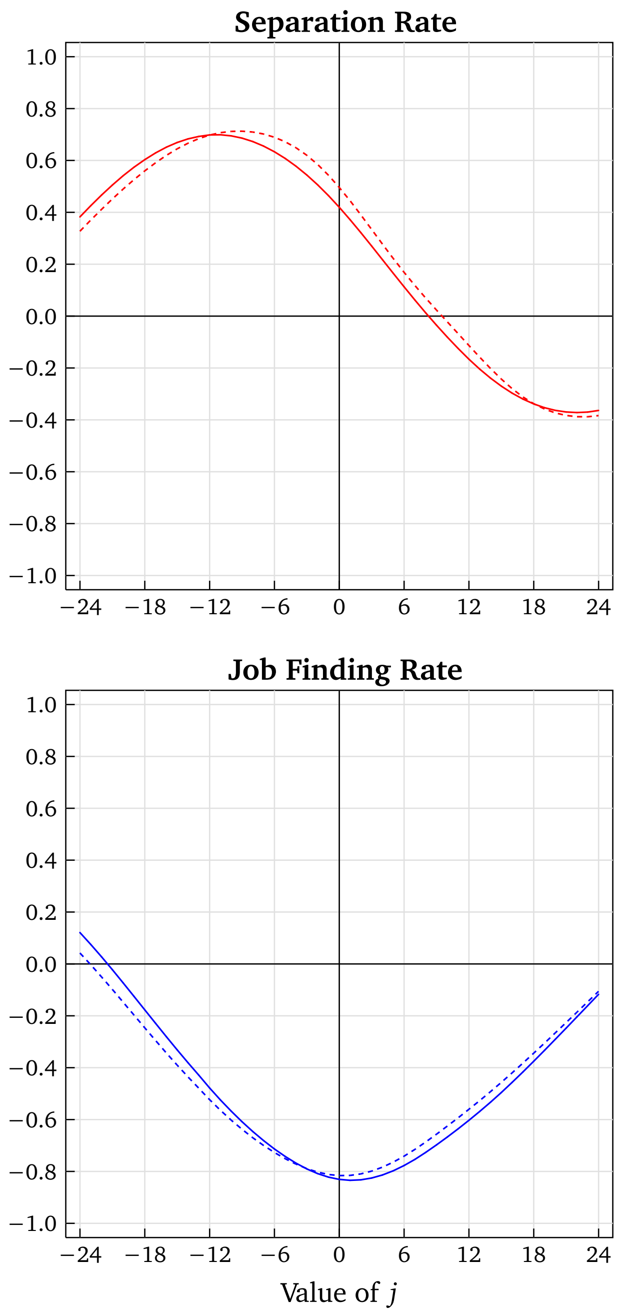

Figure 2.8. Cross-Correlations of Separation and Job Finding Hazard Rates with Unemployment, SIPP, 1983–2006[a]

Source: Author's calculations using data from the SIPP and the BLS.

a. Correlation of $\widehat{\psi}_t^{UR}$ with $\widehat{\psi}_{t+j}^{sr}$ and $\widehat{\psi}_{t+j}^{jfr}$. Cyclical component estimated using equation 2.5. Solid line is adjusted for time aggregation; dashed line is unadjusted.

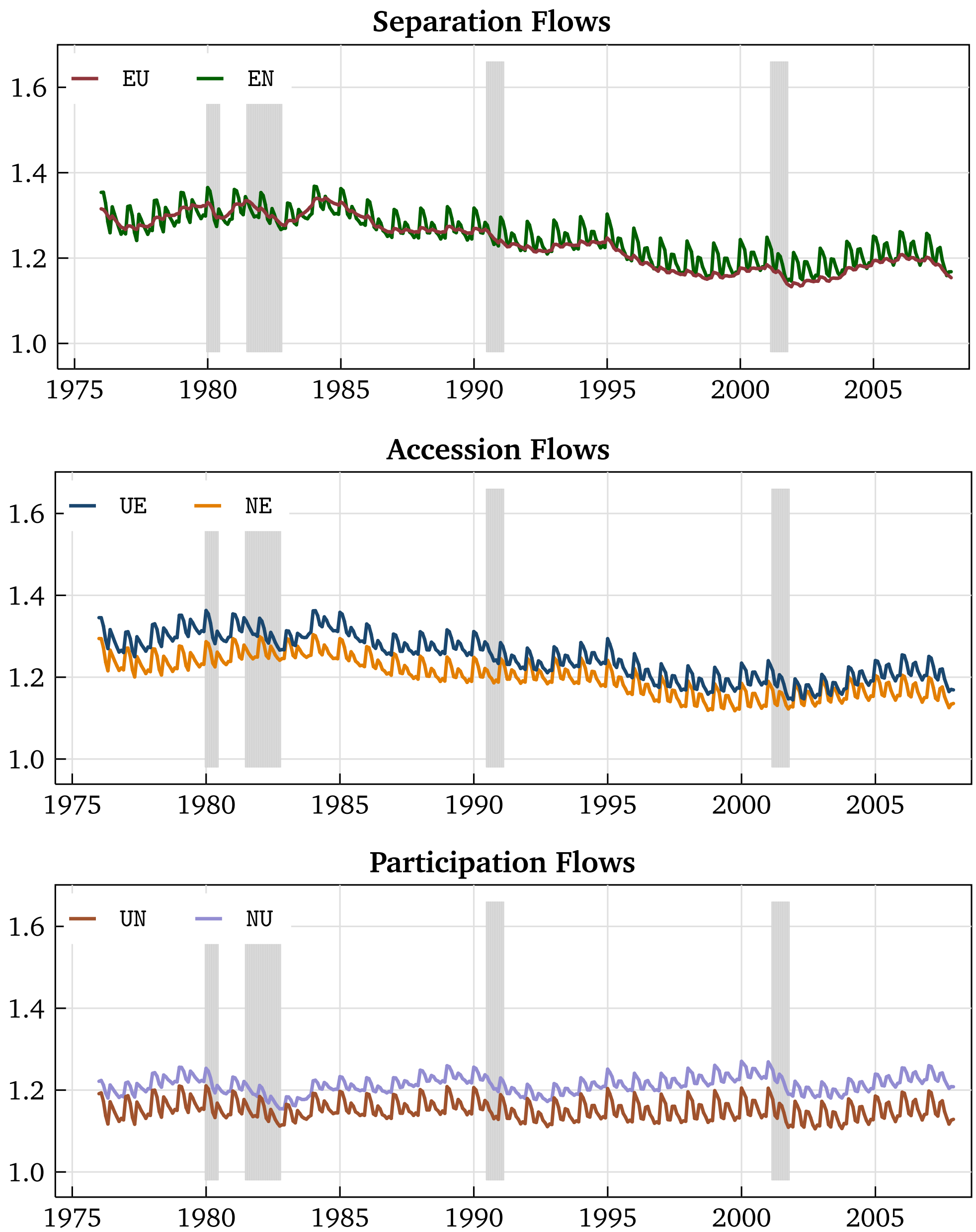

Figure 2.9. Time Aggregation Adjustment Factors, 1976–2007[a]

Source: Author's regressions using data from the SIPP and the BLS.

a. Predicted $\widehat{T}^{IJ}$ from regressions in table 2.6. Shaded bars indicate NBER-dated recessions.

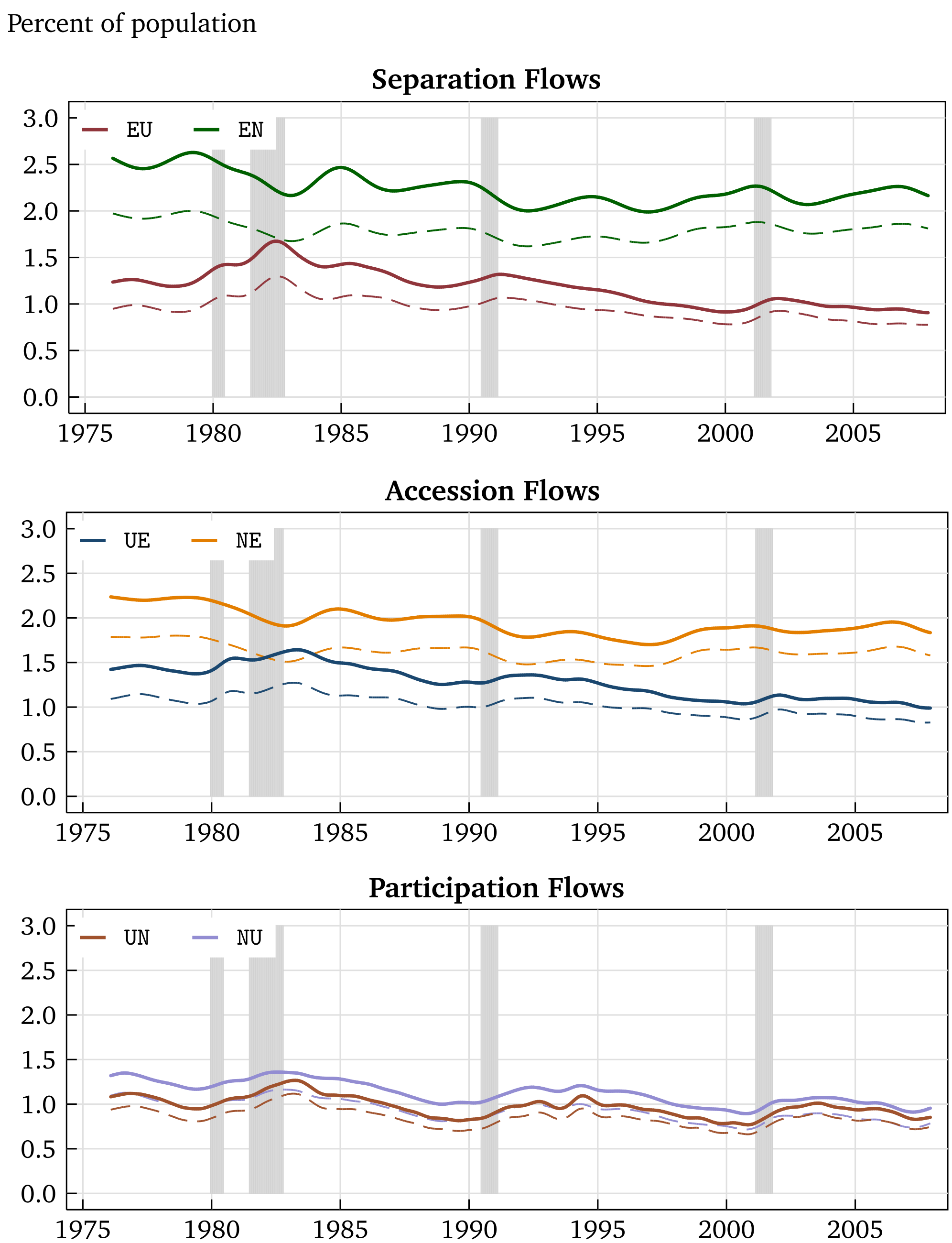

Figure 2.10. Adjusted and Unadjusted CPS Gross Flows, 1976–2007[a]

Source: Author's calculations using data from the SIPP, the CPS, and the BLS.

a. Trend-cycle component of CPS gross flows. Solid line is adjusted for time aggregation using SIPP data; dashed line is unadjusted. Cyclical component estimated using equation 2.5. Shaded bars indicate NBER-dated recessions.

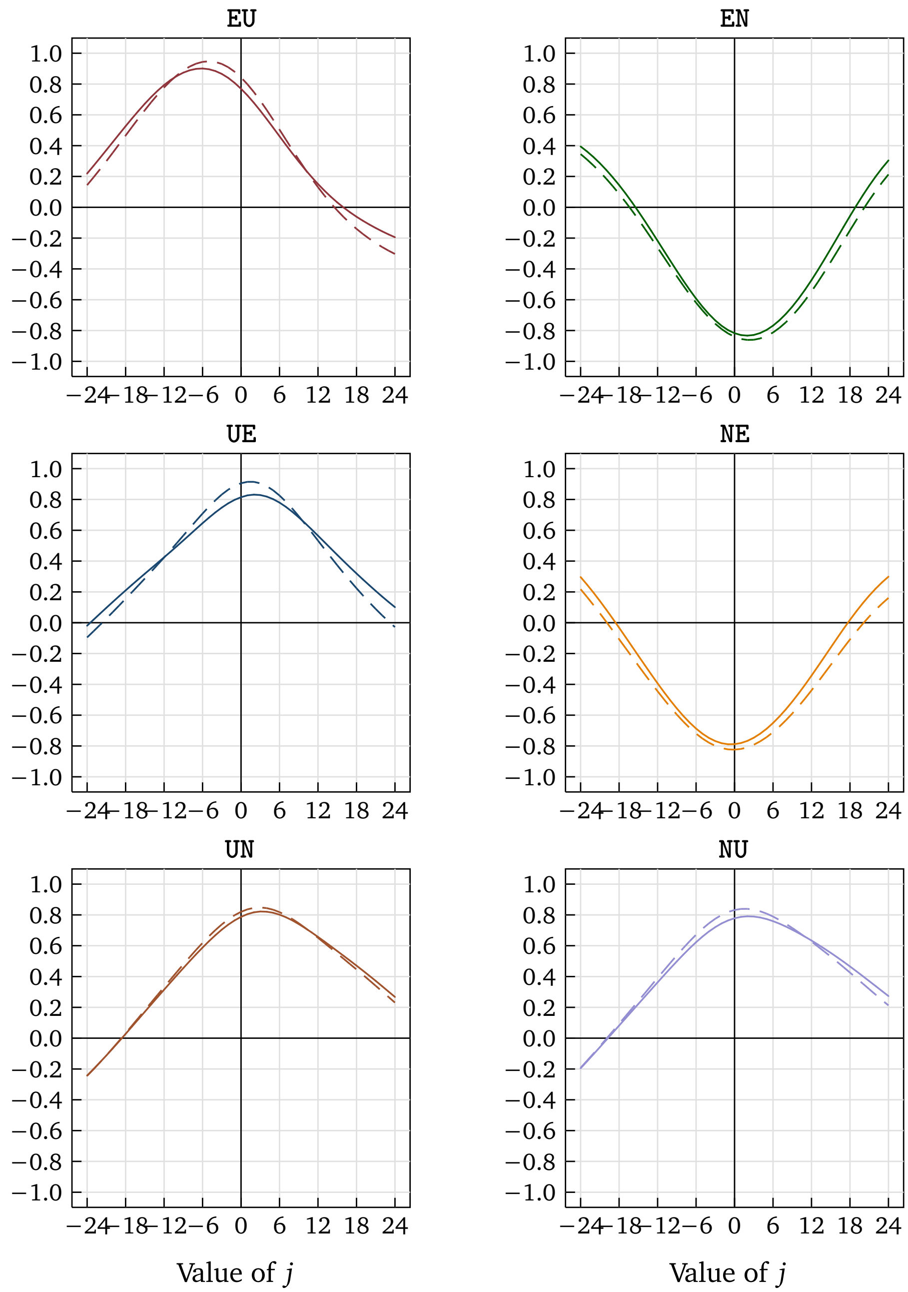

Figure 2.11. Cross-Correlations of Gross Flows with Unemployment Rate, 1976–2007[a]

Source: Author's calculations using data from the CPS and the BLS.

a. Correlation of $\widehat{\psi}_t^{UR}$ with $\widehat{\psi}_{t+j}^{\mathtt{IJ}}$. Cyclical component estimated using equation 2.5. Solid line is adjusted for time aggregation using SIPP data; dashed line is unadjusted.

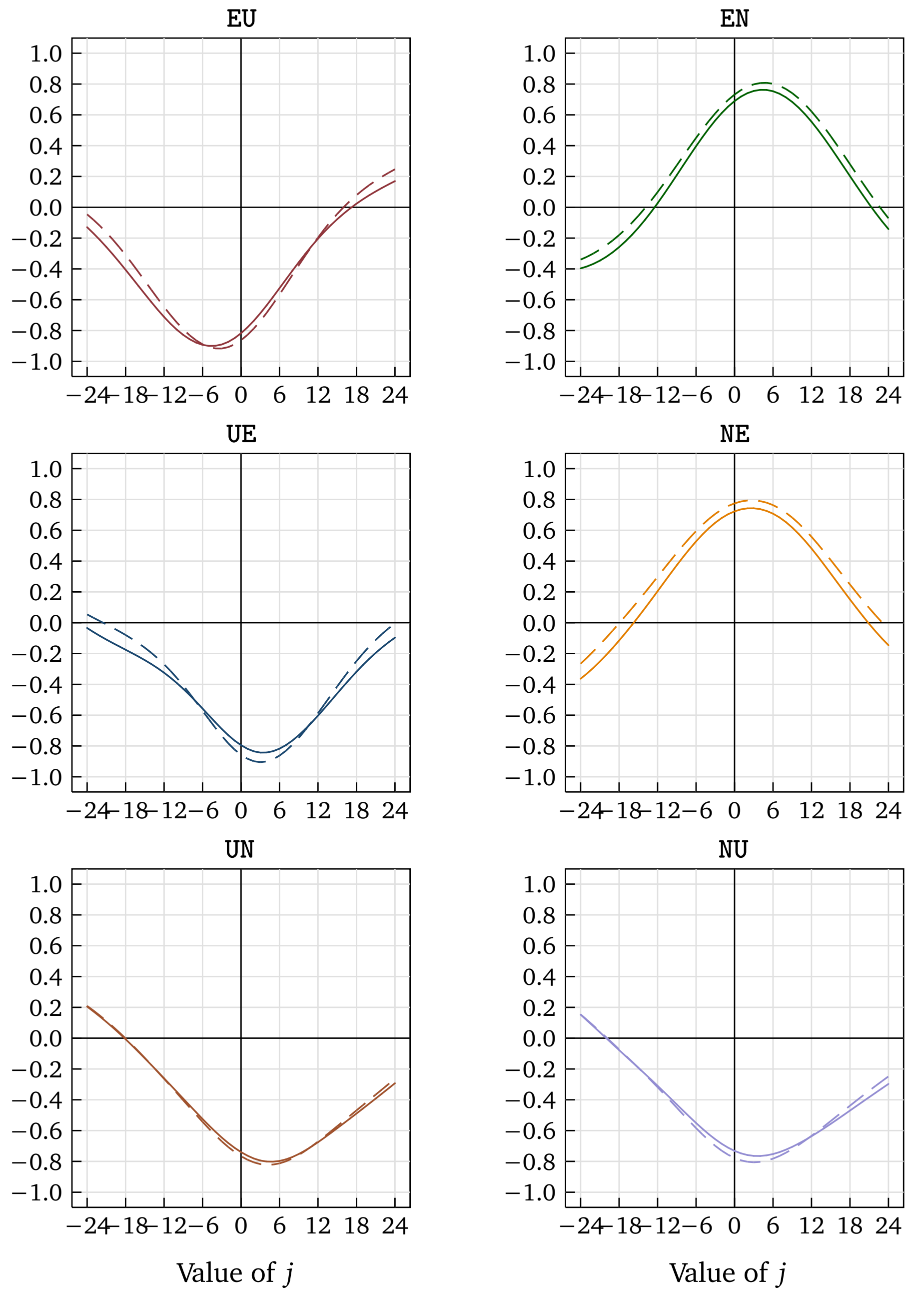

Figure 2.12. Cross-Correlations of Gross Flows with Industrial Production, 1976–2007[a]

Source: Author's calculations using data from the CPS and Board of Governors of the Federal Reserve System.

a. Correlation of $\widehat{\psi}_t^{IP}$ with $\widehat{\psi}_{t+j}^{\mathtt{IJ}}$. Cyclical component estimated using equation 2.5. Solid line is adjusted for time aggregation using SIPP data; dashed line is unadjusted.

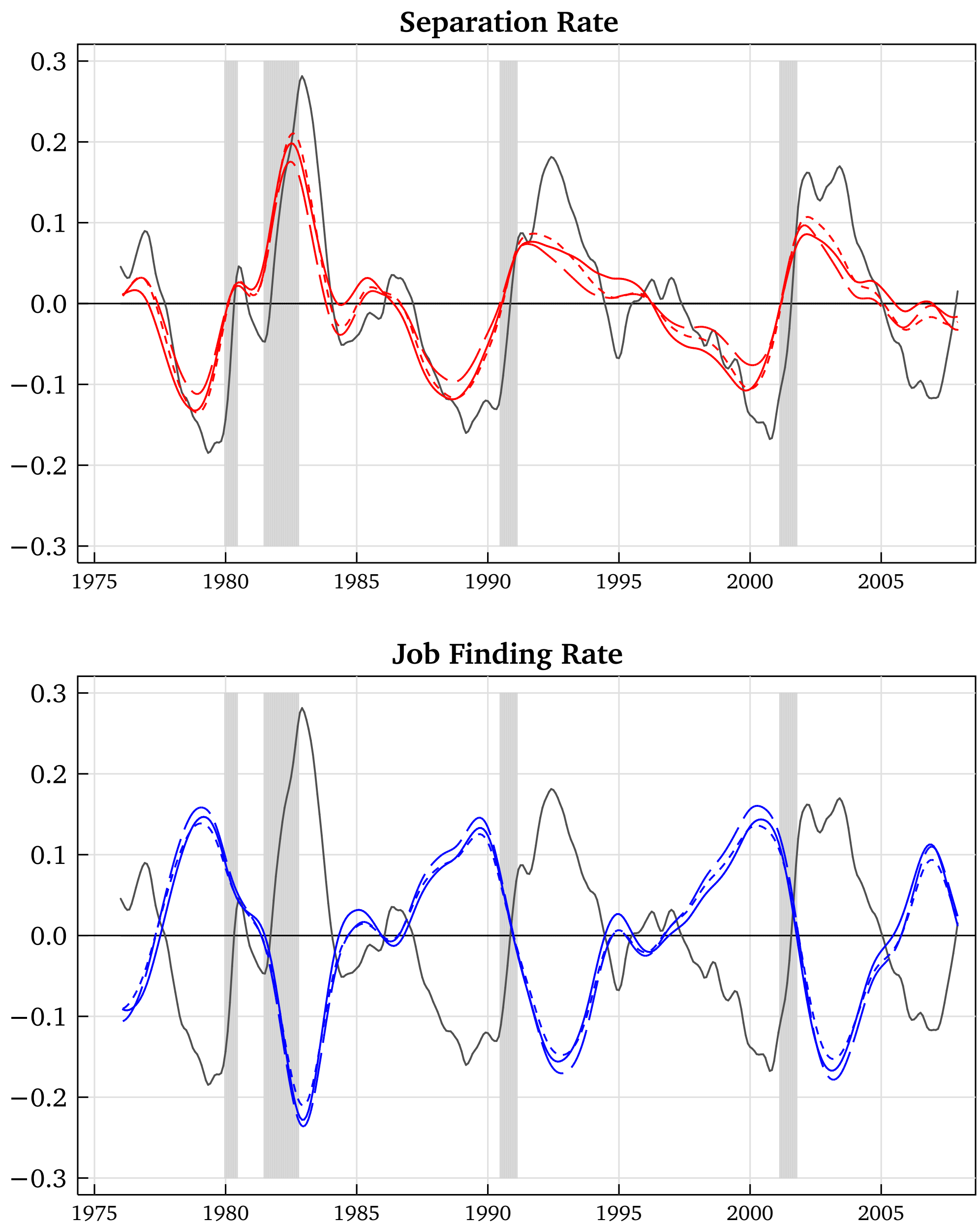

Figure 2.13. Cyclical Component of Separation and Job Finding Hazard Rates, CPS, 1976–2007[a]

Source: Author's calculations using CPS microdata for 1976:2–2007:12.

a. Cyclical component estimated using equation 2.5. Dark gray line is cyclical component of unemployment rate. Solid colored line is adjusted for time aggregation using SIPP data (equation 2.7); long-dashed line adjusted using theoretical correction (equation 2.15) and short-dashed line is unadjusted.

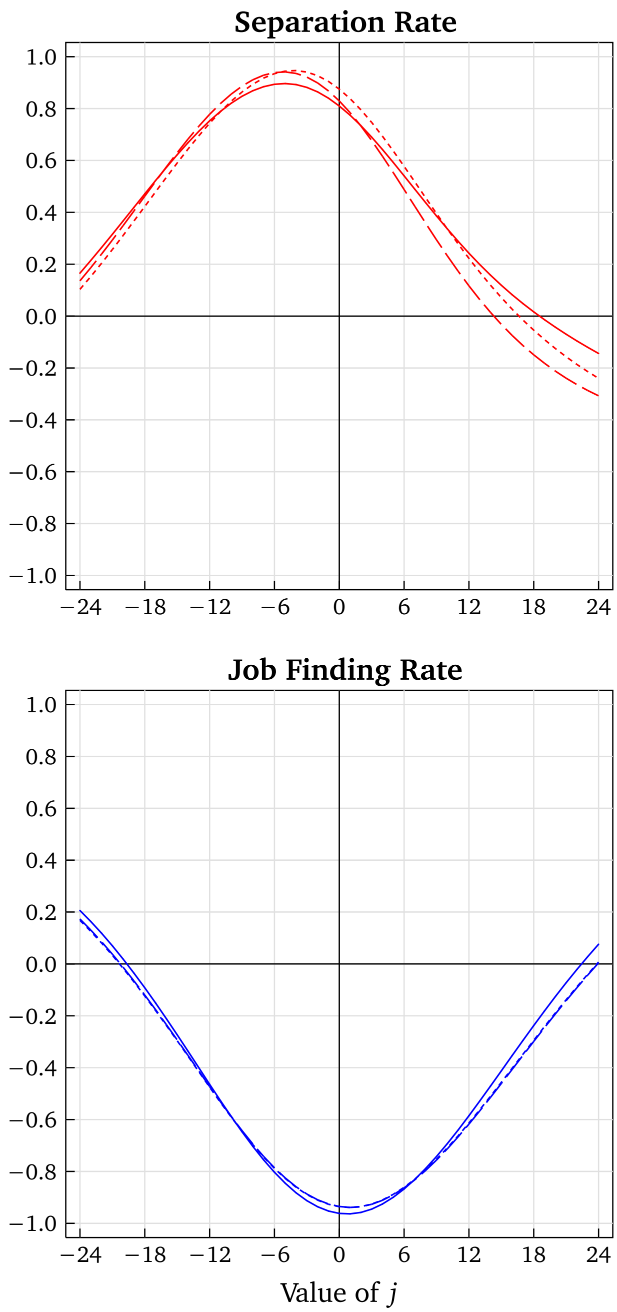

Figure 2.14. Cross-Correlations of Separation and Job Finding Hazard Rates with Unemployment, CPS, 1976–2007[a]

Source: Author's calculations using data from the SIPP, CPS, and BLS.

a. Correlation of $\widehat{\psi}_t^{UR}$ with $\widehat{\psi}_{t+j}^{sr}$ and $\widehat{\psi}_{t+j}^{jfr}$. Cyclical component estimated using equation 2.5. Solid line is adjusted for time aggregation using SIPP data (equation 2.7); long-dashed line adjusted using theoretical correction (equation 2.15) and short-dashed line is unadjusted.

Chapter 3

A Longitudinal Analysis of the Current Population Survey: Assessing the Cyclical Bias of Geographic Mobility

3.1 Introduction

Many interesting questions about the U.S. labor market are longitudinal in nature. That is, they require observations for the same individual or set of individuals at different points in time. Examples of such research are the dynamics of gross flow of workers, occupational and job mobility, the behavior of real wages over the business cycle, and the decision to migrate.

Economists generally view geographic mobility as a means of reallocating resources, in this case labor, to more efficient uses.[1] Typically 70 percent or more of people who move indicate having moved for economic reasons and up to 50 percent of those moves occurred because of a job separation.[2] In particular, researchers find a positive relationship between unemployment and geographic mobility, consistent

1. See Greenwood (1975) for a survey on mobility.
2. Lansing and Morgan (1967); Bartel (1979).

with labor reallocation.[3]

The link between labor market dynamics and mobility has important economic and public policy consequences.[4] It also has important implications for the *measurement* of labor market dynamics, particularly when using the Current Population Survey (CPS).[5] Specifically, the CPS does not follow individuals that move away from a sample address, possibly creating a bias in longitudinal measurements. Because of the strong relationship between unemployment, job separation, and mobility, there is concern that the dynamics captured by the CPS may be biased from sample attrition related to geographic mobility.

A proper assessment of this concern requires a new approach to longitudinal research using CPS data. Although the CPS is typically used in cross section, such as when calculating the unemployment rate, an individual's responses in the CPS can be matched longitudinally. Two common uses of this longitudinal feature are to match individuals from one month to the next and to match individuals from one year to the next in the CPS Annual Demographic Supplement. The longitudinal continuity allows researchers to observe changes in individuals' labor force status, income, hours worked, and many other characteristics.

Although these applications exploit the CPS's longitudinal capabilities, they do not make full use of the longitudinal information available. I create a new database that captures all possible longitudinal information in the CPS. Rather than organize the data by month, as the CPS does, I define a *person* as the fundamental unit. For each person, I combine all CPS interviews to form a mini-panel containing the largest collection of monthly observations that could possibly have come from the same person.

This database, the Longitudinal Population Database (LPD), contains the complete interview history for every person surveyed by the CPS over 1976–2006. The LPD contains data for over 10 million individuals who together are represen-

3. Bartel (1979); Schlottmann and Herzog Jr. (1981, 1984).

4. See, for example, Bartel (1979); Topel and Ward (1992); Kletzer (1998); Farber (1999); Gottschalk and Moffitt (1999); Holzer and LaLonde (1999); Neal (1999); Moscarini and Thomsson (2008).

5. Katz et al. (1984); Dahmann (1986); Welch (1993); Fitzgerald et al. (1998); Gottschalk and Moffitt (1999); Neumark and Kawaguchi (2004).

tative of the U.S. civilian noninstitutional population. Over 65 percent of these individuals have an interview history of at least four continuous months and nearly 3.3 million persons have a complete history of 8 observations.

The LPD also provides excellent information on mobility. About 20 percent of addresses in the LPD have at least one change in household. Because the LPD contains the entire history of each address in the sample, it is possible to distinguish between individuals that move ("movers") and those that do not ("stayers"). Also, since many movers spend at least four months in the sample, the LPD records their demographic characteristics and a meaningful history of labor force behavior. Furthermore, because the selection of an address for sampling is independent of the decision to move, the LPD provides a true random sample of movers. This allows a meaningful comparison of demographic and labor force characteristics of individuals who move with those who do not.

I use the LPD to assess whether geographic mobility biases labor market dynamics measured by the CPS. Comparing the populations of movers and stayers reveals minor differences in the sex, race, and education of movers but finds large differences in age and marital status of movers compared with stayers. This confirms a well-known feature of geographic mobility in the United States, the age selectivity of migration, which identifies a decline in mobility with age.[6]

Also consistent with earlier research, there are important differences in the labor force status between movers and stayers. I find that unemployed persons comprise a 60-percent greater share of the population of movers than of stayers. In addition, separations to and accessions from unemployment are twice as frequent among movers compared with stayers. Expressed as separation and job finding hazard rates, movers and stayers do not differ significantly in job finding rate but movers have a considerably higher separation hazard rate than stayers.

To assess the cyclical effect of geographic mobility, I construct a counterfactual CPS series using only the population of stayers—that is, assuming no mobility. Comparing this counterfactual series with the actual series estimated from the entire population provides a bound on the bias from geographic mobility. The bias in the

6. Gallaway (1969); Schlottmann and Herzog Jr. (1984); Tucker and Urton (1987); Peracchi and Welch (1994).

separation hazard rate moves countercyclically, implying that the separation hazard rate calculated using the entire CPS sample will appear too *acyclical*. However, the magnitude of the bias at business cycle frequencies—the difference between the cyclical component of separation hazard rates in population of stayers and in the entire sample—never exceeds 4 percent. There is little effect from geographic mobility on the job finding hazard rate.

The small cyclical bias can be reconciled with the substantial difference in separation hazard rates between movers and stayers by recognizing the distinction between individuals who move into and out of the CPS sample. The logic for a bias arising from geographic mobility bias is based on sample attrition: individuals that leave the sample are not followed. But there are equally as many people who move *into* the CPS sample as leave it and the differences between the two types of movers are small relative to stayers. Thus, the cyclical bias from geographic mobility is small because people who move out of one address tend to be replaced by similar people elsewhere in the country.

The chapter proceeds as follows. Section 3.2 briefly describes the Current Population Survey, highlighting aspects important for longitudinal matching, and explains the fundamental units of longitudinal analysis. Section 3.3 uses the LPD to assess the potential bias in the CPS due to geographic mobility. Section 3.4 explores the robustness of the bias exercise. The final section concludes.

3.2 The Longitudinal Population Database

The Current Population Survey (CPS) traces its conceptual origins back to the 1930s, when the first monthly national survey to directly measure unemployment began. The modern CPS began in 1948 as the continuation of that survey. The CPS is a monthly survey of about 50,000 U.S. households conducted to gather information about the domestic labor force. Sample households are selected at random and surveyed 8 times over sixteen months. The household rotation design was implemented to maximize continuity from month to month and year to year and to decrease the variance of survey estimates. An additional benefit of the design is that the CPS contains a wealth of longitudinal information.

Starting in the 1980s, the Census Bureau began publishing public-use microdata files containing the outcome of every CPS interview. With this information, researchers started using the CPS to explore longitudinal questions. The publicly-available CPS data are not, however, readily usable for comprehensive longitudinal research. The goal in creating the LPD is to capture all possible longitudinal information on an individual from the underlying monthly CPS surveys. The CPS is a repeated cross-section, organized by month; the LPD uses the person as the fundamental organizing unit. The LPD turns the CPS data into a panel—that is, records the complete interview history of every person surveyed. Although there is a relatively large literature about matching CPS records, previous discussions have focused on month-over-month matching.[7]

A common concern in longitudinal research using CPS data is the large number of unmatched records.[8] Roughly 30 percent of observations cannot be matched from one month to the next. Most nonmatches result from the CPS's rotating sample design, which allows at most 75 percent of individuals to match across successive months. Of observations with the potential to match, roughly 6 percent do not match—over 10 million persons a month. Viewing these missing observations in the context of their complete interview history allows missing observations to be more easily classified.

Despite these shortcomings, the CPS is an excellent survey for economic research because it is a large, random sample from the U.S. population and is the most representative sample available at this frequency. Other databases from surveys, such as the Longitudinal Research Database (LRD), the Business Employment Dynamics (BED), and the Panel Study on Income Dynamics (PSID), contain longitudinal information on their populations.[9] The LRD and BED are not ideal for studying

7. See Katz et al. (1984); Abowd and Zellner (1985); Hogue (1985); Hogue and Flaim (1986); Poterba and Summers (1984, 1986); Chua and Fuller (1987); Welch (1993); Peracchi and Welch (1994); Hausman et al. (1998); Madrian and Lefgren (2000); Shimer (2007); Moscarini and Thomsson (2008). Feng (2001) also evaluates matches using the complete interview history, but only matches the 1998 and 1999 CPS March Annual Demographic Supplement.

8. Abowd and Zellner (1985); Hogue (1985); Hogue and Flaim (1986); Welch (1993); Feng (2001); Moscarini and Thomsson (2008).

9. Dahmann (1986) discusses using panel data to study geographic mobility.

U.S. labor market dynamics because both are surveys about jobs at production establishments and not about individuals—the same person can be employed in more than one job and those not employed are not represented. In addition, they are conducted only annually or, at best, quarterly. The PSID is more appropriate for labor force research, however it is a substantially smaller sample than the CPS and is conducted only annually.

Another survey, the Job Openings and Labor Turnover Survey (JOLTS), provides monthly data on flow of hires and separations at U.S. firms. It began in December 2000 and thus provides a relative short period compared to the CPS. In addition, there are several well-documented discrepancies between aggregate estimates from JOLTS and those from other data sources.[10] In particular, the magnitude of hires and separations in JOLTS are surprisingly small compared to similar measures in other data sources.[11] More importantly, however, the JOLTS is a survey of firms, not workers. It does not include demographic information and is not suitable for studying geographic mobility.

A final U.S. household survey, the Survey of Income and Program Participation (SIPP), is also suitable for studying labor market dynamics and mobility. The SIPP is an ongoing longitudinal survey designed to study longer-term effects of income and government program participation. The SIPP panels last for between one and four years, a substantial improvement over the CPS, however the sample sizes are considerably smaller. Additionally, it is difficult to construct aggregate time-series estimates from the SIPP.[12] However, the SIPP follows individuals that move away from the initial survey address, making it ideal for studying mobility.[13]

3.2.1 Constructing the LPD

Administered jointly by the U.S. Census Bureau and the Bureau of Labor Statistics (BLS), the CPS surveys 50,000–60,000 households every month covering

10. Faberman (2005).
11. Recent work by Davis et al. (2008) devises a correction for the JOLTS data.
12. See chapter 4.
13. Neumark and Kawaguchi (2004) use the SIPP to study of how directly adjusting for geographic mobility compares to the typical Heckman (1979) selection correction.

all 50 states and the District of Columbia. It collects complete demographic and labor force data on all persons aged fifteen or older, but records basic information for all household members. Persons on active duty in the U.S. Armed Forces and persons in institutions are not eligible for survey.

The Census Bureau publishes microdata files containing the outcome of every CPS interview beginning with January 1976. I devise a complex editing and reorganization process to ensure longitudinal continuity among the different versions of the data. This section briefly describes how the LPD is constructed. Appendix D provides a detailed description and technical information.

Despite common use of the word “household,” the CPS is, in fact, a survey of *addresses*. The CPS is a multistage stratified sample of addresses from 792 sample areas in the United States.[14] Housing units are sampled from address lists generated from the Decennial Census of Population and Housing and updated for housing built after the census. The sample is drawn once per decade using information from the most recent Decennial Census.

The CPS uses a rotating sample to minimize variance, both between months and between households, as well as reduce the burden on respondents. Each address selected for the sample is surveyed for four consecutive months, not surveyed for the next eight months, and then surveyed again for the next four months. It then leaves the sample permanently.

An address is identified over time by its month in sample (MIS) designation, which corresponds to the number of times the address is scheduled to be surveyed. Figure 3.1 shows the relationship between the MIS and the calendar month of the survey and also the sample rotation. The 4-8-4 rotation pattern enables up to 75 percent of units to match from one month to the next and 50 percent to match from year to year. The large continuity between households across time permits sophisticated longitudinal analysis using CPS data.

The first time an address enters the sample it is visited in person by a Census Bureau field representative to establish whether it is eligible for survey. To be considered eligible the housing unit at the address must be occupied by at least one

14. Bureau of Labor Statistics (2002), chapter 3.

person eligible for interview (a civilian who is at least fifteen years old and does not usually reside elsewhere). At eligible housing units, the surveyor initiates the CPS interview.

Ineligible addresses are recorded as a noninterview. A *type C noninterview* occurs if the address is permanently ineligible for interview. This condition arises if the housing unit has been converted to a permanent business, condemned, or demolished or if the address falls outside the area for which it was selected. The address is never visited again. A *type B noninterview* occurs if the address is intended for occupancy but is not occupied by any eligible person. Such units are typically vacant, but also include those occupied entirely by individuals not eligible for interview. Type B addresses may become eligible in the future and are thus visited for all eight months that the address is in the sample.

The previous two types of noninterview occur when no one from the civilian noninstitutional population resides at the selected address. Such locations are not considered part of the CPS sample. The third type, a *type A noninterview*, occurs if the address is eligible for a CPS interview but no useable data are collected. This can arise because the occupants are absent or otherwise unavailable during the interviewing period or refuse to participate in the interview. These noninterviews are considered part of the CPS sample. However, because no information about the current occupants is collected, the sample weight of similar nearby units is increased to compensate. The type A condition is considered temporary and the address is visited in all succeeding months.

The BLS assigns each household a scrambled identifier to ensure confidentiality but still permit longitudinal matching. For data after 1994, when the CPS was substantially redesigned, the household identifier is globally unique. Prior to 1994, however, it is only unique across two months for households in the same rotation group. I develop an algorithm to identify households and generate a globally-unique household identifier.[15]

In addition, the BLS periodically changes the scrambled identifier for households. This is disruptive for longitudinal matching in the LPD. For simple month-

15. Feng (2001) develops a similar procedure to exploit the pattern of sample rotation.

over-month matching, a change of household identifier prohibits a match only for the month in which the change occurred; all preceding and subsequent months match. However, because the LPD matches an individual across sixteen months, an identifier change disrupts longitudinal continuity for the entire history. Authors either report a missing value for the month where matching was impossible or construct a moving-average across months that do match.

A second challenge in constructing the LPD is ensuring longitudinal consistency. Over the thirty-year period for which microdata are available, the data definitions change 17 times. I develop a consistent set of definitions for categorical variables (e. g., race, educational attainment, or occupation) for the entire LPD.

After creating longitudinally-consistent variables and unique household identifiers for every month, the data are combined together to form the LPD. The LPD has over 53 million observations covering the period 1976–2007, or approximately 140,000 observations per month. The smallest month has just under 97,000 observations and the biggest month almost 160,000.

3.2.2 Longitudinal Units of the LPD

The objective of the LPD is to construct a complete longitudinal record for every person in the CPS. The CPS, however, is a probability sample of addresses, not individuals. Therefore, constructing a person's longitudinal history begins with the interview history at the address level. In any month, an address is occupied by a single household. But households can move into and out of an address during its time in the sample, generating a difference between the household and the address. Each household consists of one or more individuals. As with addresses, individuals may move into and out of a household. Thus each individual must be identified longitudinally in relation to her household and address. Figure 3.2 shows the hierarchical relationship between addresses, households, and persons.

An *interview history* is the collection of all monthly observations from a particular unit (address, household, or person). The address is the basic unit. All households and persons from an address inherit the same address interview history. The household is subset of the address. All individuals within a household share the same household interview history but each household has a unique interview history. The finest unit is the person. Each person has a unique interview history. Table 3.1 provides example interview histories for different longitudinal situations encountered in the CPS. This table will be referenced throughout the following subsections.

Addresses

Each sample address is scheduled for 8 interviews by a Census Bureau field representative. An *address observation unit* (AOU) is the collection of interviews conducted at an address during its time in the CPS sample. An AOU can have at most 8 observations, but addresses found permanently ineligible (type C noninterviews) will have fewer than 8. Many type C noninterviews are determined on the first interview or following a type B noninterview. Example 3 in table 3.1 shows the interview history for an address with a type C noninterview. There are about 3.7 million unique AOUs in the LPD (table 3.2).

Households

Because an AOU spans sixteen months, including eight months without being surveyed, it is possible for more than one household to occupy the address during its time in the CPS. Households that move during the survey are not followed by the CPS; instead the replacement household, if any, is surveyed for the rest of that address's time in the sample. A *household observation unit* (HOU) is the largest collection of observations within an AOU that can possibly come from the same household.

Because individuals are identified within their household, AOUs must first be examined to identify unique households. In most cases an AOU contains only one household, but some AOUs have at least one change of household. A household change can occur in 4 ways:

H1. The original occupants of the address move out and a replacement household moves in with no intervening vacancy recorded.

H2. The original occupants of the address move out and a replacement household moves in but with an intervening vacancy.

H3. The original occupants of the address move out and are not replaced during the address's tenure in the CPS sample.

H4. The address is initially vacant but a household moves in before the address has rotated out of the sample.

The household change in case H1 is straightforward. The replacement household is identified as a household by the CPS. There is no noninterview recorded in the AOU, however it must be partitioned into two HOUs to reflect the change in household. Individuals associated with the original HOU are replaced by the new occupants. Example 2 in table 3.1 demonstrates such a situation.

In case H2 the replacement household is often not identified as part of a new household. Accordingly, the LPD creates a separate HOU any time a string of completed interviews within an AOU is interrupted by one or more type B noninterviews. Example 5 in table 3.1 depicts such a situation. The AOU contains a type B noninterview at MIS 4. The first HOU within this AOU contains the first three observations; the remaining completed interviews are assigned to the second HOU.

When a previously-occupied housing unit is found ineligible during all remaining months in the CPS (case H3), the subsequent type B noninterviews are discarded. This is depicted in table 3.1, example 6. Similarly, when an address is found initially ineligible but subsequently interviewed (case H4) the initial type B observations are discarded. Line 7 of table 3.1 shows an example of case H4 and the resulting HOUs. Both cases, however, identify households that have moved.

Over 80 percent of AOUs in the LPD have no household change (table 3.2). These households are known not to have moved during their tenure in the CPS. This does *not*, however, imply that these HOUs have no noninterviews. Type A noninterviews are permitted within an HOU and do not imply mobility. The remaining addresses, just over 19 percent, record a change of household during their tenure in the CPS.

Household changes interrupt longitudinal continuity of the observation unit.

For research where continuity is important, such as calculating gross labor force flows, these interruptions reduce the number of observable transitions. For other avenues of research, however, these household changes are beneficial. In particular, a change within an AOU identifies a household that has moved.

On average, each address is occupied by 1.14 households over its sixteen months in the CPS sample. The implied rate of annual mobility, the probability that a household does not reside at the same address one year later, is 14.7 percent.[16] This rate is consistent with the annual rate of geographic mobility estimated by the Census Bureau using the CPS Annual Demographic Supplement. U.S. Census Bureau (2007) reports the average annual mobility rate over 1976–2007 is 14.9 percent.

Persons

Each household has one or more persons residing there. A *person observation unit* (POU) is the largest collection of observations within an HOU that can possibly come from the same person. Because the POU is a subset of the HOU, all POUs within that HOU also terminate when an HOU ends. Example 2 in table 3.1 demonstrates this: the POU for the person in household 1 terminates when the second household begins.

Also, because individuals can move into and out of a household, each POU can have a different interview history from its associated HOU. Consider, for example, a college student living with her parents during summer: she is counted in the household for interviews conducted during the summer, but her POU terminates when she returns to school. Such a case is shown for the 2nd person in example 8 (table 3.1).

There are 10.6 million unique POUs in the LPD. The CPS collects full demographic and labor force information only for persons over fifteen years old. For those younger than fifteen, only information on sex, race, and age is collected. There are 2.3 million POUs for persons aged fifteen years and younger. These POUs are not included when studying mobility.

16. The LPD contains 4,160,835 unique households at 3,646,370 unique addresses, yielding 1.1380 households per address (table 3.2). This implies an annual rate of mobility equal to $1-(1.1380/16\times 12) = 0.1465$.

3.2.3 Longitudinal Statistics from the LPD

How useful the LPD is depends on how much meaningful longitudinal information is contained within the POUs. This section provides a detailed analysis of the POUs and reveals the large amount of longitudinal information contained in the LPD.

For each POU I calculate the number of attempted interviews and the number of completed interviews. For example, the individuals in example 1 from table 3.1 both have 8 attempted interviews and 8 completed interviews. In example 2, person 1 from household 1 has 4 attempted interviews and 4 completed interviews. The individual in example 3 has 5 attempted interviews, all completed.

Table 3.3 reports tabulations of the number of attempted and completed interviews for all POUs. POUs are weighted by the average CPS sampling weight for the POU.[17] The column totals (bottom) are the share of POUs with that number of completed interviews. The row totals (right) are the share of POUs with that number of attempted interviews. Thus cells on the diagonal are POUs with no noninterviews; these contain the most longitudinal information possible. The sum of the diagonal elements, the share of POUs without missing observations, makes up 94 percent of the LPD.

Because a POU combines two blocks of consecutive monthly interviews, it is also important to identify the number of consecutive months of longitudinal information. The bottom right cell shows POUs with 8 completed interviews, that is, two four-month blocks. It is also the single largest cell, accounting for 31 percent of all POUs.

But many more POUs have at least one block of four months. The next largest cell in table 3.3 is for 4 completed interviews out of 4 interviews, comprising 26 percent of POUs. Persons with a block of four interviews are important for studying mobility, because often the other block is missing because of a move. The LPD contains about 5.8 million POUs, just under 60 percent of all POUs, with either 4 or 8 completed interviews and no noninterviews.

17. See section 3.2.4 for details.

3.2.4 Match Validity

The standard procedure in the literature is to match observations from one month to the next using household and person identification variables and then validate these matches using supplementary demographic characteristics.[18] A failure of any criterion invalidates the match.

The LPD allows for much more sophisticated evaluation of matched observations. Instead of evaluating the match just from one month to the next, the entire interview history can be used. I develop a measure that evaluates each month against all other months for a person, rather than simply month-over-month.

For example, consider a man who is mistakenly classified as a woman for one month of his tenure in the CPS. The standard validation procedure would potentially discard 2 matches (1/3 of the total possible) from this simple mistake (one match on either side of the classification error). One failed match criterion over 8 observations on a person, is very likely to be a clerical mistake and not an invalid match. My method evaluates each month using *all* longitudinal information for the person. In particular, responses for each month are evaluated against those in all other months.

I evaluate a match's validity according to 3 criteria

1. Sex: a person's sex should not change over the POU.
2. Race: a person's race should not change over the POU.
3. Age: a person's age should not change by more than 2 years over the POU.

To formalize, let s_{it} indicate the sex recorded for person i in month t. Similarly, let r_{it} and a_{it} be the recorded race and age in month t. Person i has T_i valid observations in the LPD. The validity score V_{it} of the month t observation for person i is

$$V_{it} = \frac{1}{3T_i}\sum_{j=1}^{T_i} I(s_{it} = s_{ij}) + I(r_{it} = r_{ij}) + I(|a_{it} - a_{ij}| \leq 2), \tag{3.1}$$

where $I(\cdot)$ an indicator function that is 1 if the statement is true and 0 otherwise. For a person with only one observation, $V_{it} = 1$.

18. For example, Madrian and Lefgren (2000) consider sex, race, age, and educational attainment. Shimer (2007) and Moscarini and Thomsson (2008) use sex, race, and age.

If all criteria match for all observations, $V_{it} = 1$ for all t. In the example above, each month's score falls because of the failure of the sex criterion. However in month where sex was female, the score is lower still, because $I(s_{it} = s_{ij}) = 0$ for all other months. Thus, this method penalizes all of a person's observations for a single failure; the month with the discrepancy is penalized more.

I treat V_{it} as representing the probability of valid match and adjust the person's month t sampling weight, ω_{it}, by that probability to get the validity-weighted sampling weight $\nu_{it} = \omega_{it} V_{it}$. All population estimates are calculated using this adjusted sampling weight. Thus, each labor force transition is effectively weighted by the "probability" that it came from the same person.[19]

The average validity score in the LPD is 0.9604 when taken over all observations and 0.9930 when taken over nonmissing observations (those with positive CPS sampling weight). This confirms that most matches directly identified by the CPS are valid. In addition, since only the latter group enter population totals, the observed match quality is very high. The results that follow are robust to using other match validation procedures; see section 3.4.1.

3.3 Geographic Mobility

Geographic mobility has important implications for the *measurement* of labor market dynamics, particularly when using the CPS. Specifically, the CPS does not follow individuals that move away from a sample address, possibly creating a bias in longitudinal measurements. Because of the strong relationship between unemployment, job separation, and mobility, there is concern that the dynamics captured by the CPS may be biased from sample attrition related to geographic mobility.

The argument that geographic mobility can bias longitudinal measurements is usually phrased in terms of sample attrition: some event, possibly related to the business cycle, causes a household to move out of the CPS sample. Therefore, because the CPS does not follow those individuals that leave the sample, there may be a cyclical bias from geographic mobility.

19. Feng (2001) evaluates the probability of a valid match conditional on sex, race, age, and marital status using Bayes' rule. This still, however, leads to a binary accept-reject decision.

Sample attrition is not, however, the only type of mobility observed. As section 3.2.2 emphasized, a change of household at an address can occur 4 different ways, only 1 of which (H3) is pure sample attrition. In fact, the LPD identifies roughly equal numbers of persons moving into and out of the sample. Thus, the language of "sample attrition" is not the correct way to describe geographic mobility in the CPS. Instead, I describe mobility in terms of "out-movers" and "in-movers". An *out-mover* is a person who permanently leaves an address during its tenure in the CPS sample. An *in-mover* is a person not originally present who joins at an address during its tenure in the sample.

3.3.1 Identifying Geographic Mobility

Geographic mobility is identified using the interview histories in the LPD. Using the full longitudinal history of a person allows me to identify persons that move separately from persons with missing observations arising from some other reason. Two types of mobility can be identified. The first is a complete change of household. This is the most common, accounting for 70 percent of movers.

The population of movers is identified as the set of observations for which the interview history of the HOU differs from that of the AOU. This definition captures all mobility events described by cases H1–H4 and combinations thereof. Mobility is not identified simply based on the number of observations in an HOU nor the existence of missing observations. Instead, mobility is identified using the LPD by the relationship between the HOU and its AOU. For example, line 9 of table 3.1 shows a case where no interview was recorded in MIS 7. However because the interview history for the AOU and HOU are identical, this household is considered a nonmover. All AOUs with at least one valid observation that have a type B noninterview are identified as movers.

In addition to households that move, individuals can move into and out of households. Examples 8 and 9 in table 3.1 show such cases. Individuals that move into and out of an HOU are *not* included in the population of stayers. Individual mobility—that is, not associated with a household change—accounts for 23 percent of movers. The remaining 6 percent combine both household and individual mobility.

Figure 3.3 shows the distribution of completed interviews per POU, decomposed into the contribution by stayers and movers. Each bar represents the share of total of POUs with N completed interviews; its height is a graphical representation of the bottom row of table 3.3. Within each category, the bottom segment of the bar represents the share of total POUs that came from stayers while the top 2 segments represent the contribution of movers.

Of POUs with 4 or fewer completed observations, movers account for 55 percent of the total. The share of movers drops substantially for those with 5 to 8 completed interviews, accounting for 27 percent on average; movers' share declines monotonically to zero. In-movers account for almost three-quarters of POUs with 1-3 completed interviews. There are about the same number of in-movers and out-movers with 4 completed interviews.

The significant decrease in the share of movers with more than 4 completed interviews is sensible. Even if the probability of moving stays constant, the greatest likelihood of observing a move lies in the 8 months when the person is not in the sample. This predicts a substantial, discrete fall in the share of movers after the first group of four months.

3.3.2 Demographic Characteristics of Movers and Stayers

Before assessing the bias from geographic mobility, this section examines characteristics of the populations of stayers and movers. If the population of persons that move is similar to those that do not, then their movement into and out of the CPS sample will cause little bias. However if the population of movers differs substantially from those who do not move, the bias from mobility may be large. In addition, it is important to distinguish between in-movers and out-movers. Even if movers differ from stayers, if persons who move into the CPS sample resemble those who leave it, then the bias from mobility may be small.

Table 3.4 reports the population proportions for several demographic characteristics.[20] The first column shows the proportion of all persons in the LPD with the

20. The populations are calculated using validity-weighted sampling weights; see section 3.2.4 for

indicated characteristic. The second column reports the proportion for stayers and the third and forth columns report the proportions for out-movers and in-movers.

The population of movers does not differ significantly in sex from those that do not move. Also, the populations of in-movers and out-movers have nearly the same ratio of females to males as the population of stayers. The other demographic characteristics have more meaningful differences. There are more nonwhite movers than stayers: the population of stayers is 85.1 percent white, compared with 81.8 percent for movers. Roughly 60 percent of the difference is accounted for by black movers. In-movers and out-movers do not differ appreciably in race.

A well-known feature of geographic mobility is the so-called "age selectivity of migration," which identifies a decline in mobility with age.[21] To assess this difference I classify age into 3 functional groups: younger (sixteen to twenty-four), prime age (twenty-five to fifty-four), and older (fifty-five and older). Table 3.4 confirms the age selectivity of migration: movers are younger than stayers. The population of movers has twice as many persons aged sixteen to twenty-four compared to stayers. Again, the difference between in-movers and out-movers is not large. The proportion of prime-age movers is basically the same as for stayers, implying an equally dramatic difference in the share of those aged fifty-five and older. The proportion of older movers is less than one-half that of stayers. Because prime-age workers are more likely to be in the labor force relative to those younger or older, the relative homogeneity in this category may mitigate potential bias from geographic mobility.

There are almost 80 percent more persons who have never married in the population of movers compared with stayers. Those never married account for 37 percent of movers but only 22 percent of stayers. The share of widowed and divorced are nearly identical between movers and stayers, implying that married persons are significantly less likely to move. The proportion married is 63 percent among stayers compared to 46 percent for movers.

There is relatively little difference in education between movers and stayers. The bottom panel of table 3.4 reports the distribution of educational attainment,

---

details.

21. Gallaway (1969); Schlottmann and Herzog Jr. (1984); Tucker and Urton (1987); Peracchi and Welch (1994).

divided into 4 functional categories: less than a high school education, high school graduates, some college, and college graduates. Movers are slightly more likely to be high school drop-outs or in college.

Although there is little or no difference in the distribution of movers and stayers by sex, race, and education, there are large differences in age and marital status. Individuals who move are more likely to be nonwhite, young, not married, and in college. In addition, because movers represent roughly 25 percent of all POUs, these differences will be economically meaningful if the characteristics are correlated with labor force status.

3.3.3 Labor Force Characteristics of Movers and Stayers

There are clear differences in the demographic characteristics of individuals who move and those who do not. This section explores whether those differences are also reflected in labor force status and transitions. As before, the first column of table 3.5 reports the proportion of all persons in the LPD with the indicated characteristic, the second column reports the proportion for stayers, and the third and forth columns report the proportions for out-movers and in-movers.

There are substantial differences in the distribution of labor force status between the movers and stayers (top panel, table 3.5). The population of movers has about one-fifth as many persons not in the labor force (NILF) and correspondingly more employed and unemployed. In particular, there are twice as many more unemployed movers than stayers. Unlike with demographics, there are significant differences in labor force status between in-movers and out-movers.[22] There are about 7 percent more unemployed out-movers than in-movers, suggesting a link between job loss and mobility.

The lower panel of table 3.5 reports the population proportions for labor force transitions. Nontransitions, that is a "transition" between the same labor force state, are not reported.[23] The bottom 3 rows show unobserved transitions: transitions for which the previous month's labor force status is not known. These repre-

22. For similar findings see Bartel (1979); Schlottmann and Herzog Jr. (1981, 1984).
23. Nontransitions account for 66 percent of all transitions and 93 percent of *observed* transitions.

sent a substantial fraction of all transitions (30 percent). The discrepancy between measured stocks and gross flows that arises because of these missing transitions is known as "margin error."[24]

The first row shows that separations to unemployment (EU transitions) account for 0.62 percent of all labor force transitions in the CPS over 1976–2007. Among movers, however, EU transitions account for 0.93 percent of transitions, over 70 percent more than among stayers. Similarly, UE transitions account for an 80-percent larger share of mover's transitions than stayers'. Transitions between employment and nonparticipation also occur with greater frequency among movers, but the differences are more modest.

Most missing observations arise because of the CPS's rotating sample design, which ensures that at most 75 percent of the sample matches from one month to the next. However, unmatched observations also occur because of type A noninterviews, clerical errors, and mobility. Movers will have a greater share of missing observations because out-movers are not followed and because the history of in-movers is unknown. In particular, because transitions are defined with respect to the current month, there are more missing transitions for in-movers than for out-movers.

This is confirmed in the bottom 3 rows of table 3.5, where movers have higher population proportions than stayers. In-movers record roughly 20 percent more missing transitions than do out-movers. A truly striking result is that transitions from missing to unemployment (XU) are almost three times as prevalent for movers. In contrast, transitions to employment are "only" 40 percent higher among movers. An important implication of these findings is that margin error–adjustment should be calculated separately for movers and stayers.[25]

24. See Abowd and Zellner (1985); Poterba and Summers (1984, 1986); Chua and Fuller (1987); Fujita and Ramey (2006).
25. See appendix C.

3.3.4 Bias from Geographic Mobility

The population of individuals who move is different from those who do not.[26] Although the LPD identifies individuals that move and contains information on those persons while they are in the sample, it does not, of course, say anything about them when they are not in the sample.

Because the CPS does not follow households that move, estimates of movers' gross flows and hazard rates from the LPD may not accurately reflect that population's true behavior. However it is possible to conduct the counterfactual experiment of what the CPS data would show if there was no mobility by considering only the population of stayers. Comparing this counterfactual series with the actual series estimated from the entire population provides a bound on the bias from geographic mobility.

Let σ_t be the share of the month t population that does not move:

$$\sigma_t = \frac{\mathtt{P}_t^S}{\mathtt{P}_t}, \tag{3.2}$$

where $\mathtt{P}_t$ is the total population and superscript S denotes stayers. The average of σ_t over 1976–2007 is 0.7798. This value is not strictly comparable to the estimates of mobility rates presented earlier, which measure the number of persons not living at the same address one year later.

The total number of persons who transition from state $\mathtt{I}$ in month $t-1$ to state $\mathtt{J}$ in month t can be divided into the number of transitions made by stayers and those by movers:

$$\mathtt{IJ}_t = \mathtt{IJ}_t^S + \mathtt{IJ}_t^M, \tag{3.3}$$

where superscript M denotes movers. This implies a similar decomposition of the separation and job finding hazard rates:

$$s_t = \sigma_t s_t^S + (1-\sigma_t)s_t^M \quad \text{and} \quad f_t = \sigma_t f_t^S + (1-\sigma_t)f_t^M, \tag{3.4}$$

where s_t and f_t are the separation and job finding hazard rates for the entire CPS sample.

26. Whether these observed differences are the ex ante cause of mobility or the ex post result of mobility is a separate and interesting question.

The monthly separation and job finding hazard rates are calculated by

$$\widehat{s}_t = \frac{\mathtt{EU}_t}{\mathtt{E}_{t-1}} \quad \text{and} \quad \widehat{f}_t = \frac{\mathtt{UE}_t}{\mathtt{U}_{t-1}} \tag{3.5}$$

for the entire CPS population and by

$$\widehat{s}_t^S = \frac{\mathtt{EU}_t^S}{\mathtt{E}_{t-1}^S} \quad \text{and} \quad \widehat{f}_t^S = \frac{\mathtt{UE}_t^S}{\mathtt{U}_{t-1}^S} \tag{3.6}$$

for the population of stayers, where $\mathtt{E}$ and $\mathtt{U}$ are the stock of employed and unemployed persons.

A way to assess the potential bias from geographic mobility is to measure the difference between the hazard rate calculated for stayers and the entire CPS sample. Define the ratio between the counterfactual hazard rate and the measured hazard rate as

$$G(s)_t = \frac{\widehat{s}_t^S}{\widehat{s}_t} \qquad \text{and} \qquad G(f)_t = \frac{\widehat{f}_t^S}{\widehat{f}_t} \tag{3.7}$$

If the hazard rates of the populations of movers and stayers are identical this ratio is 1; $G \neq 1$ indicates differences attributable to geographic mobility.

The upper panel of table 3.6 reports the averages of $G(s)$ and $G(f)$ over 1976–2007. The average ratio of the job finding hazard rate of movers to that of stayers is nearly 1, indicating that the job finding hazard rate of stayers does not differ much from that of the whole population. The average job finding hazard rate is about 2 percent lower for stayers than for the entire population.

In contrast, the separation hazard rate of stayers is almost 20 percent lower than that for the entire population. This implies that the separation rate for movers is much higher than in the total population. Indeed, the separation rate calculated from the available information from the population of movers is 65 percent higher than that from the entire sample. This value should be interpreted cautiously, however, because some labor force behavior of movers is not observable.

Nevertheless, there is a clear difference between movers and stayers in their probability of separating to unemployment: movers have a substantially higher separation hazard rate. Movers and stayers to do not differ significantly in job finding behavior, however. Although the effect on the level of separations is large, of principal concern is whether geographic mobility affects the *cyclical* behavior of hazard rates. If the difference between the separation rate of stayers and the general population does not change significantly over the business cycle, geographic mobility contributes little bias.

3.3.5 Cyclical Bias

I model the observed time series as the sum of four independent, unobserved components: a trend, a cycle, a seasonal, and an irregular component.[27] The trend represents low-frequency movements in the series. The cyclical component is a stochastic periodic function of time with a frequency at that of the business cycle. The seasonal component represents fluctuations that repeat annually and the irregular component captures the remaining non-systematic variation.

The structural time series model for the natural logarithm of each series, denoted y_t, is

$$y_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t\,, \tag{3.8}$$

where μ_t is the trend, ψ_t the cyclical, γ_t the seasonal, and ε_t the irregular component. Details of the econometric specification of the components are provided in appendix B.

Equation 3.8 is recast as a state space model where the unobserved components are represented by the state of the system. The unknown parameters are estimated by maximum likelihood using the Kalman filter to update and smooth the unobserved state. The estimation is performed using the STAMP program written by Koopman et al. (2007). See appendix B for details.

A reasonable concern is that mobility associated with the business cycle may create a cyclical bias in measured gross flows and hazard rates. First, however, it is important to understand how mobility changes over the business cycle. I measure the annual rate of geographic mobility by one minus the share of persons reported living at the same address one year later reported by the U.S. Census Bureau.[28]

27. This follows the general method described in Harvey (1989).
28. See U.S. Census Bureau (2007).

I isolate the cyclical component of mobility by estimating equation 3.8 at annual frequency.[29] Figure 3.4 plots the cyclical component of the mobility rate together with that of the unemployment rate for comparison.[30] The cyclical component of mobility tends to follow the unemployment rate, indicating that more people move during recessions than during booms. This is consistent geographic mobility as a means for reallocating idle labor to more productive uses. The contemporaneous correlation of the cyclical component of the mobility rate with the unemployment rate is 0.50, confirming the apparently countercyclicality. The peak correlation of 0.51 trails unemployment by two months.

I next estimate equation 3.8 for each of the four hazard rates in equations 3.5 and 3.6. I evaluate the cyclical bias using the ratio measure G (equation 3.7). In this case the ratio is calculated as the difference in the log cyclical components:

$$G(\psi^s)_t = \widehat{\psi}^{s,S}_t - \widehat{\psi}^s_t \qquad \text{and} \qquad G(\psi^f)_t = \widehat{\psi}^{f,S}_t - \widehat{\psi}^f_t, \tag{3.9}$$

where ψ^s and ψ^f are the cyclical components of the separation and job finding hazard rates calculated from the whole population and $\psi^{s,S}$ and $\psi^{f,S}$ are those calculated from only the population of stayers.

Summary statistics for $G(\psi^s)$ and $G(\psi^f)$ over 1976–2007 are reported in the lower panel of table 3.6. The values in the lower panel are the percentage difference between the cyclical component of the hazard rate calculated from the population of stayers and the hazard rate calculated from the entire population. The minimum and maximum values indicate the greatest degree of bias from geographic mobility. The cyclical dynamics of job finding are not effected by mobility; the largest cyclical difference is 1 percent. There is a more modest effect of mobility on the separation hazard rate, although the peak bias never exceeds 4 percent.

Figures 3.5 and 3.6 plot the cyclical components of the actual and counterfactual hazard rate series (ψ_t and ψ^S_t). The cyclical component of the unemployment rate (in gray) is also shown for comparison. The solid line plots the hazard rate for the entire population while the dashed line uses only the population of stayers. The

29. This eliminates the seasonal component.

30. For the graph only, I use a locally weighted polynomial regression smoother (Cleveland, 1979) to create a monthly time series of the cyclical component from the annual data.

lower panel shows the estimated cyclical bias, $G(\psi)_t$. The vertical axes are drawn so the divisions of the left and right ordinates have the same size.

The separation hazard rate of stayers, shown in figure 3.5, is more volatile at business cycle frequencies than the separation hazard rate of the entire population. It generally falls further at the cyclical peak (the trough of unemployment) and rises higher at the cyclical trough (the peak of unemployment). The cyclical bias, shown in the lower panel of figure 3.5, reflects this pattern. The cyclical correlation of the bias with the unemployment rate (table 3.7) is 0.55, indicating moderate countercyclicality. That is, the bias from geographic mobility rises during recession as more people move.

Figure 3.6 shows that there is little effect of geographic mobility on job finding hazard rates. The hazard rate calculated from the population of stayers is largely indistinguishable from that calculated using the entire population. Although the bias from geographic mobility, shown in the lower panel of figure 3.6, is mildly procyclical (table 3.7), the difference between job finding hazard rates measured from stayers and the whole population never exceeds 1 percent.

3.3.6 Discussion

The LPD allows me to identify individuals who move into and out of the CPS sample. Because many movers spend four months or more in the sample, I can observe their demographic characteristics and establish a meaningful history of labor market behavior. Comparing the populations of movers and stayers reveals no difference in the composition of sex and minor differences in race and education. There are, however, large differences in age and marital status of movers compared with stayers.

There are also substantial differences in the distribution of labor force status between the two populations: there are 60 percent more unemployed movers than unemployed stayers. In addition, `EU` separations and `UE` accessions comprise almost twice the share of transitions for movers than for stayers. Separations and accessions are best interpreted in the context of separation and job finding hazard rates. Movers have a substantially higher separation hazard rate than stayers, although they do not differ significantly in job finding rate.

Geographic mobility varies negatively with the business cycle, possibly creating a cyclical bias to measured separation and job finding hazard rates. The bias in hazard rates arising from not observing the behavior of movers can be assessed by comparing a counterfactual hazard rate calculated from the population of stayers to the hazard rate calculated for the entire population.

The cyclical bias in the separation hazard rate is countercyclical, meaning that the separation hazard rate calculated using the entire CPS sample will appear too *acyclical*. There is little effect of geographic mobility on the job finding hazard rate.

This evidence can be interpreted as follows. The rate of separations to unemployment and of geographic mobility both increase during a recession. The separation hazard rate of stayers rises more during a recession than does the entire sample, implying that the separation rate of movers is *less* countercyclical.[31] Put differently, during a boom the separation hazard rate falls, however the separation rate of movers falls by less than the entire population. Nevertheless, the cyclical difference between the separation hazard rate of stayers and the entire population never exceeds 4 percent; geographic mobility does not significantly affect the cyclicality of measured hazard rates.

This relatively small bias seems at odds with the substantial differences in average separation hazard rates between movers and stayers (table 3.6). These differences can be reconciled by recognizing the importance of differentiating between out-movers and in-movers. The argument of geographic mobility bias is one of sample attrition: a person leaves the sample and is not followed. But focusing solely on out-movers is misguided. There are equally as many in-movers as out-movers (by person) and in-movers account for 60 percent of movers' observations.

In addition, the demographic and labor force evidence presented in tables 3.4 and 3.5 shows that, although movers are quite different from stayers, the differences

31. This is confirmed by estimating the cyclical component of movers' separation hazard rate. The cyclical correlation with unemployment is 0.85, compared with 0.88 in the entire sample. As before, this relationship should be interpreted cautiously because the full history of the population of movers is not observed.

between in-movers and out-movers are small, especially relative to stayers. Thus, appealing to the CPS's random sampling, a person who moves out of one address is replaced by a similar in-mover elsewhere in the country and the true bias from sample attrition (i. e., out-movers) is offset by similar in-movers.

3.4 Robustness

This section examines the robustness of the analysis of geographic mobility. I consider alternative measures of validating matches of observations in the LPD and assess my findings using an alternate procedure for isolating cyclical components.

3.4.1 Alternate Measures of Match Validity

This section evaluates match validity using 2 alternative criteria. The first criterion is "naive" matching, that is matches determined solely by the information that defines which observations can match. A second criterion is to consider the average validity score for a person,

$$\overline{V}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} V_{it}, \tag{3.10}$$

where V_{it} is defined in equation 3.1, and use a threshold rule to determine which matches are counted. I included all persons where $\overline{V}_i \geq 0.875$. In practice, this value allows for 1 failure among the 3 criteria over a four month block of observations.

I then calculate the separation and job finding hazard rates under each of the alternate criteria. To facilitate comparison across the 3 criteria, the hazard rates are expressed relative to the baseline of weighted matching. Table 3.8 reports the summary statistics for these two measures. There is virtually no difference in the measures. Naive matching yields almost identical results as probability-weighted matching. There are larger differences when using the threshold criterion, but the effects are still quite modest. Both separation and job finding hazard rates are slightly lower, with a peak difference of about 4 percent.

There are two central results. First, overall match quality in the LPD is very high. The average validity score over all observations in the LPD is 0.9930. This is not a feature of the LPD per se, but of the underlying CPS data. The second is not surprising given the first: adjusting for matches supplemental validity does not significantly affect results.

3.4.2 Alternate Method of Isolating Cyclical Component

In this section I explore an alternate method for isolating the cyclical component of the time series. A common technique in macroeconomics is to filter the seasonally-adjusted series using the Hodrick-Prescott (HP) filter to extract the cyclical component.[32]

Because the HP filter requires a continuous time series, any missing observations associated with changes in the household identifier must be interpolated. Researchers use either a local moving average or linear interpolation to create a continuous time series. For simplicity, I use linear interpolation. I next seasonally adjust the series using the Census Bureau's X-12-ARIMA seasonal adjustment program. Finally, I HP filter the seasonally-adjusted series with smoothing parameter $\lambda = 129,600$.[33]

Figures 3.7 and 3.8 plot the cyclical components of the separation and job finding hazard rate using the alternative cyclical isolation procedure. The cyclical components are considerably more volatile, particularly at high frequencies. This high-frequency volatility is a natural consequence of the HP filter, which removes only the low-frequency trend. As such, it is more difficult to clearly identify cyclical patterns from the graph of the time series than in figures 3.5 and 3.6, particularly distinguishing between the actual and counterfactual series.

As the lower panels in figures 3.7 and 3.8 show, the cyclical bias estimated using the HP filter is considerably larger than that estimated using the unobserved-components model. The bias from geographic mobility contributes up to 15 percent for the separation rate and 10 percent for the job finding rate (table 3.9).

Although the degree bias is larger, its correlation with unemployment weakens dramatically when using the HP filter (table 3.10). Although the signs remain

32. Hodrick and Prescott (1997).
33. Ravn and Uhlig (2002) find this to be the optimal smoothing parameter for monthly data.

the same, the correlations in the HP data are essentially acyclical. The correlation of the bias in the job finding hazard rate with unemployment is not statistically significant at the 10 percent level. Even though the degree of bias is large, it is unrelated to the business cycle identified by the HP filter.

Note also that the cyclical correlation of the hazard rates fall when using the HP filter to isolate the cyclical component. Using data for the entire population, the cyclical correlation with unemployment for the separation hazard rate falls from 0.87 to 0.60 and from -0.94 to -0.73 for the job finding hazard rate. Nevertheless the HP filter shows a strong relationship of both hazard rates with the business cycle.

3.5 Conclusion

Because the CPS does not follow individuals that move away from a sample address, the strong relationship between unemployment, job separation, and mobility creates concern that labor market dynamics captured by the CPS may be biased from sample attrition related to geographic mobility. Using a new database that permits sophisticated longitudinal analysis of the all CPS data, I find that the cyclical bias arising from geographic mobility is small. At business cycle frequencies, the difference between the separation hazard rate calculated from the entire CPS sample and from a subset that is known not to have moved never exceeds 4 percent. There is little effect from geographic mobility on the job finding hazard rate.

To facilitate this study of mobility and of other important longitudinal research topics, I construct a new database, the Longitudinal Population Database (LPD), that organizes the CPS data into individual panels, where the person is the fundamental unit. I develop a novel framework for identifying an individual's full longitudinal history inside a survey that is, fundamentally, a sample of addresses. The LPD, contains the complete interview history for every person surveyed by the CPS over 1976–2006, over 10 million individuals. Over 65 percent of persons have a interview history of at least four continuous months and nearly 3.3 million have a complete history of 8 observations.

The LPD provides excellent information on mobility. Because the LPD contains the entire history of each address in the sample, it is possible to distinguish between movers and stayers and between in-movers and out-movers. About 25 percent of individuals in the LPD move at some point during their tenure in the sample. Since many movers spend at least four months in the sample, the LPD records their demographic characteristics and a meaningful history of labor force behavior. Furthermore, because the selection of an address for sampling is independent of the decision to move, the LPD contains a true random sample of movers.

Comparing the populations of movers and stayers reveals only minor differences in the composition of sex, race, and education of movers and stayers. However, I find that movers are younger than stayers and more movers are unmarried. Movers are also more likely to be unemployed.

I assess labor market dynamics using the separation and job finding hazard rates. On average, the separation hazard rate of stayers is almost 20 percent lower than that for the entire population, implying a high separation rate for movers. The separation rate of movers is, indeed, about 65 percent higher than when using the entire population. There is relatively little difference in the job finding hazard rates between movers and stayers.

This large difference in average separation hazard rates between movers and stayers seems at odds with the small degree of cyclical bias. This tension can be reconciled by distinguishing between out-movers and in-movers. The argument of geographic mobility bias is one of sample attrition: a person leaves the sample and is not followed. But focusing solely on out-movers is misguided because there are equally as many in-movers as out-movers. In addition, the demographic and labor force evidence shows that the differences between in-movers and out-movers are small relative to those between movers and stayers.

The evidence presented in this chapter is consistent with the idea that geographic mobility reflects efficient resource reallocation. Geographic mobility increases during a recession, facilitating the reallocation of idle resources—unemployed persons—across space to more productive uses. Fortunately, this labor reallocation does not significantly impact the measurement of U.S. labor market dynamics.

Table 3.1. Examples of Relationship among Addresses, Households, and Persons in the LPD[a]

	Interview history		
Example	*Address*	*Household*	*Person*
1. No noninterviews, no household change	iiii iiii	iiii iiii	iiii iiii
2. No noninterviews, household change	iiii iiii	iiii ____	iiii ____
	iiii iiii	____ iiii	____ iiii
3. Type C noninterview	iiii iC__	iiii i___	iiii i___
4. All noninterviews	BBBB BBC_	____ ____	____ ____
5. Intervening type B noninterview	iiiB iiii	iii_ ____	iii_ ____
	iiiB iiii	____ iiii	____ iiii
6. Trailing type B noninterview(s)	iiii iBBB	iiii i___	iiii i___
7. Initial type B noninterview(s)	BBBB iiii	____ iiii	____ iiii
8. An individual out-mover	iiii iiii	iiii iiii	iiii iiii
	iiii iiii	iiii iiii	iii_ ____
9. An individual in-mover	iiii iAii	iiii iAii	iiii iAii
	iiii iAii	iiii iAii	___i iAii

a. An i denotes a completed interview; A, B, and C denote type A, type B, and type C noninterviews; and a underscore denotes a missing observation. Each row per example depicts a separate household or person.

Table 3.2. Total Number of Longitudinal Units[a]

Unit	*Stayers*	*Movers*	*Total*
Addresses (AOU)	2,948,860	707,510	3,656,370
	[80.6]	[19.4]	[100.0]
Households (HOU)	2,959,414	1,201,421	4,160,835
	[71.1]	[28.9]	[100.0]
Persons[b] (POU)	7,984,300	2,617,606	10,601,906
	[75.3]	[24.7]	[100.0]

Source: Author's calculations using LPD data for 1976:1–2007:12.

a. Numbers in brackets are percent of total units.

b. Includes persons under sixteen years old.

Table 3.3. Number of Completed Interviews per POU, by Number of Attempted Interviews

Percent[a]

Attempted interviews	*Completed interviews*								
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*Total*
1	10.7								10.7
2	0.4	7.0							7.5
3	0.1	0.5	7.5						8.1
4	0.2	0.4	1.4	25.3					27.4
5	0.1	0.0	0.1	0.4	2.7				3.4
6	0.0	0.0	0.0	0.1	0.2	2.5			2.9
7	0.0	0.0	0.1	0.1	0.1	0.3	3.9		4.5
8	0.2	0.1	0.2	0.5	0.4	0.7	2.6	30.9	35.6
Total	11.7	8.1	9.3	26.5	3.4	3.5	6.5	30.9	100.0

Source: Author's calculations using LPD data for 1976:1–2007:12.

a. Share of validity-weighted count of person observation units (POUs) for persons aged sixteen years or older.

Table 3.4. Distribution of Demographic Characteristics, by Mobility

Percent[a]

Category	*LPD*	*Stayer*	*Out-mover*	*In-mover*
Sex				
Male	48.2	47.4	49.8	49.5
Female	51.9	52.6	50.2	50.5
Race				
White	84.3	85.7	82.2	81.1
Black	11.8	10.8	13.3	13.8
Other	4.0	3.5	4.6	5.1
Age				
16–24	21.2	15.5	31.9	32.9
24–54	54.9	55.0	53.2	55.4
55 and older	23.9	29.5	14.7	11.7
Marital status				
Married	57.6	63.2	45.7	46.8
Widowed or divorced	14.8	14.8	15.4	14.6
Never married	27.6	22.0	39.0	38.7
Education				
High school drop-out	23.9	23.6	24.6	24.6
High school graduate	35.8	35.8	35.2	35.9
Some college	21.3	21.0	22.4	21.8
College graduate	19.0	19.6	17.8	17.6

Source: Author's calculations using LPD data for 1976:1–2007:12.

a. Calculated using validity-weighted sampling weight for persons aged sixteen years or older.

Table 3.5. Distribution of Labor Force Characteristics, by Mobility

Percent[a]

Category	*LPD*	*Stayer*	*Out-mover*	*In-mover*
Labor force status				
E	61.8	61.2	63.4	64.6
U	4.0	3.3	6.3	6.8
N	34.2	35.6	30.3	28.7
Labor force transition				
EU	0.6	0.5	1.0	0.9
EN	1.3	1.2	1.6	1.3
UE	0.7	0.6	1.1	1.1
NE	1.1	1.1	1.2	1.2
UN	0.6	0.5	0.9	0.8
NU	0.6	0.5	0.9	0.8
XE	19.0	17.5	21.4	26.3
XU	1.4	1.0	2.3	3.2
XN	10.2	10.0	10.1	12.0

Source: Author's calculations using LPD data for 1976:1–2007:12.

a. Calculated using validity-weighted sampling weight for persons aged sixteen years or older.

Table 3.6. Effect of Geographic Mobility on Hazard Rates[a]

Hazard rate	*No. of obs.*	*Mean*	*Standard deviation*	*Minimum*	*Maximum*
Level[b]					
Separation	369	0.8048	0.0505	0.6593	0.9512
Job finding	369	0.9809	0.0335	0.8847	1.0779
Cyclical component[c]					
Separation	382	−0.0002	0.0143	−0.0382	0.0331
Job finding	382	0.0001	0.0037	−0.0075	0.0102

Source: Author's calculations using LPD data for 1976:1–2007:12.

a. Calculated using validity-weighted sampling weight for persons aged sixteen years or older.

b. Ratio of hazard rate calculated from population of stayers to that of entire population; see equation 3.7.

c. Difference between cyclical component estimated from population of stayers and that of entire population; see equation 3.9.

Table 3.7. Contemporaneous Cyclical Correlation with Unemployment[a]

Hazard rate	*All*	*Stayers*	*Bias*[b]
Separation	0.8743	0.8813	0.5537
Job finding	−0.9396	−0.9379	−0.1957

Source: Author's calculations using LPD data for 1976:1–2007:12.

a. corr $\left(\psi_t^{UR}, \psi_t^{y}\right)$, where y is the item listed in the column head.

b. Difference between cyclical component estimated from population of stayers and that of entire population; see equation 3.9.

Table 3.8. Separation and Job Finding Hazard Rates, Alternate Measures[a]

Ratio to validity-weighted measure

Measure	*No. of observations*	*Mean*	*Standard deviation*	*Minimum*	*Maximum*
Separation					
Naive	369	1.0000	0.0000	1.0000	1.0000
Threshold	366	0.9939	0.0081	0.9666	1.0182
Job finding					
Naive	369	1.0000	0.0000	1.0000	1.0000
Threshold	366	0.9995	0.0057	0.9755	1.0129

Source: Author's calculations using LPD data for 1976:1–2007:12.

a. Naive does not validate matches; threshold keeps all matches with $\overline{V}_i \geq 0.875$.

Table 3.9. Effect of Geographic Mobility on Hazard Rates, Alternate Measure[a]

Variable	*No. of observations*	*Mean*	*Standard deviation*	*Minimum*	*Maximum*
Level[b]					
Separation	382	0.8078	0.0524	0.6830	0.9684
Job finding	382	0.9820	0.0320	0.8642	1.0898
Cyclical component[c]					
Separation	382	0.0000	0.0522	−0.1391	0.1582
Job finding	382	0.0000	0.0316	−0.1203	0.1044

Source: Author's calculations using LPD data for 1976:1–2007:12.

a. Calculated using validity-weighted sampling weight for persons aged sixteen and older.

b. Ratio of hazard rate calculated from population of stayers to that of entire population; see equation 3.7.

c. Difference between cyclical component estimated from population of stayers and that of entire population; see equation 3.9. Cyclical component isolated from seasonally-adjusted series using Hodrick-Prescott filter.

Table 3.10. Contemporaneous Cyclical Correlation with Unemployment, Alternate Measure[a]

Hazard rate	*All*	*Stayers*	*Bias*[b]
Separation	0.6042	0.5817	0.1180
Job finding	−0.7255	−0.6909	−0.0464

Source: Author's calculations using LPD data for 1976:1–2007:12.

a. $\operatorname{corr}\left(\psi_t^{UR}, \psi_t^{y}\right)$, where y is the item listed in the column head. Cyclical component isolated from seasonally-adjusted series using Hodrick-Prescott filter.

b. Difference between cyclical component estimated from population of stayers and that of entire population; see equation 3.9.

	Month																			
PID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
A	1	2	3	4									5	6	7	8				
B		1	2	3	4									5	6	7	8			
C			1	2	3	4									5	6	7	8		
D				1	2	3	4									5	6	7	8	
E					1	2	3	4									5	6	7	8
F						1	2	3	4									5	6	7
G							1	2	3	4									5	6
H								1	2	3	4									5
I									1	2	3	4								
J										1	2	3	4							
K											1	2	3	4						
L												1	2	3	4					
M													1	2	3	4				
N														1	2	3	4			
O															1	2	3	4		
P																1	2	3	4	
Q																	1	2	3	4

Figure 3.1. CPS Sample Rotation Pattern

Address Observation Unit

The collection of all interviews conducted at an address

Removed

AOUs with all type B or C noninterviews

+ Includes Type A and Type B noninterviews
+ Type C noninterviews end an AOU

Household Observation Unit

The largest collection of observations within an AOU that could possibly come from the same household

Type B noninterviews between households

+ A Type B or Type C noninterview ends an HOU
+ Any remaining observations form a new HOU
+ A replacement household starts a new HOU

Person Observation Unit

The largest collection of observations within a HOU that could possibly come from the same person

Nothing

+ Initial type A noninterviews remain with the person succeeding them
+ Intervening type A noninterviews do not start a new POU
+ Trailing type A noninterviews remain with the person preceding them

Figure 3.2. Hierarchy of Observation Units

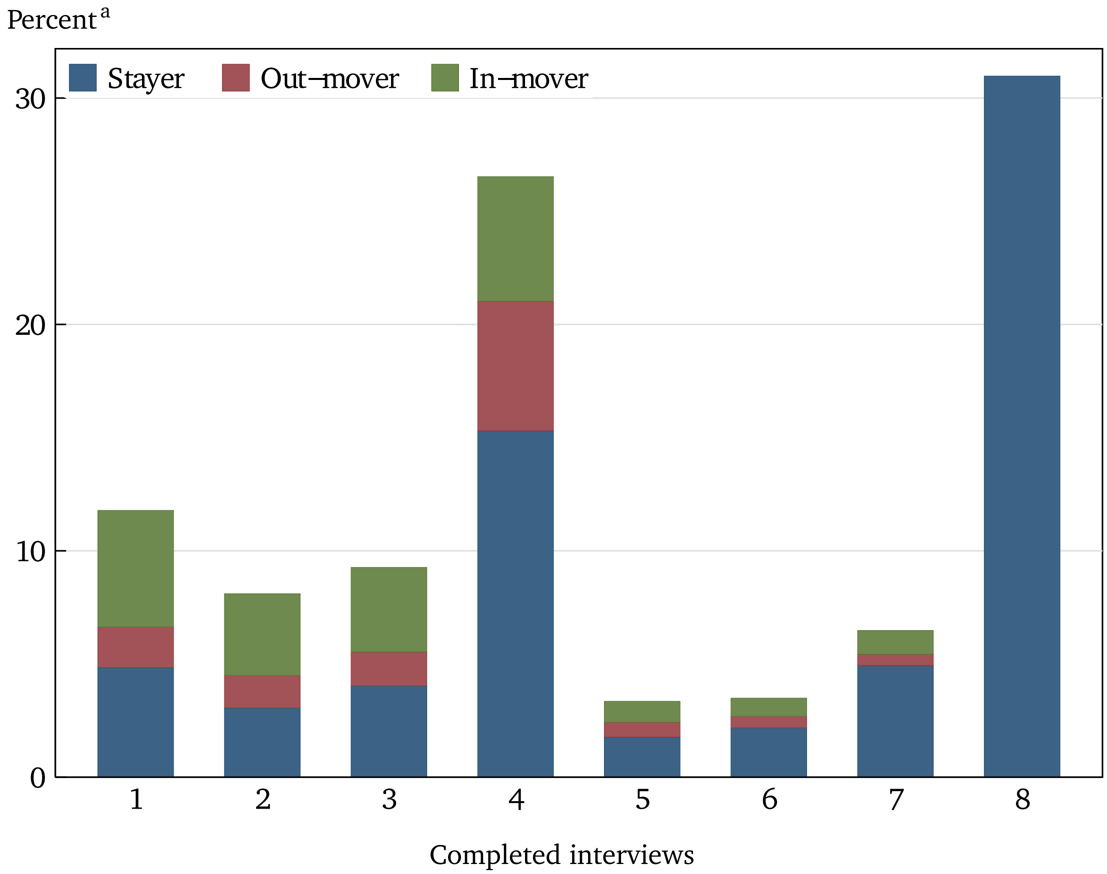

Figure 3.3. Distribution of Completed Interviews per POU, by Mobility

Source: Author's calculations using LPD microdata for 1976:1–2007:12.

a. Share of validity-weighted count of person observation units (POUs) for persons aged sixteen years or older.

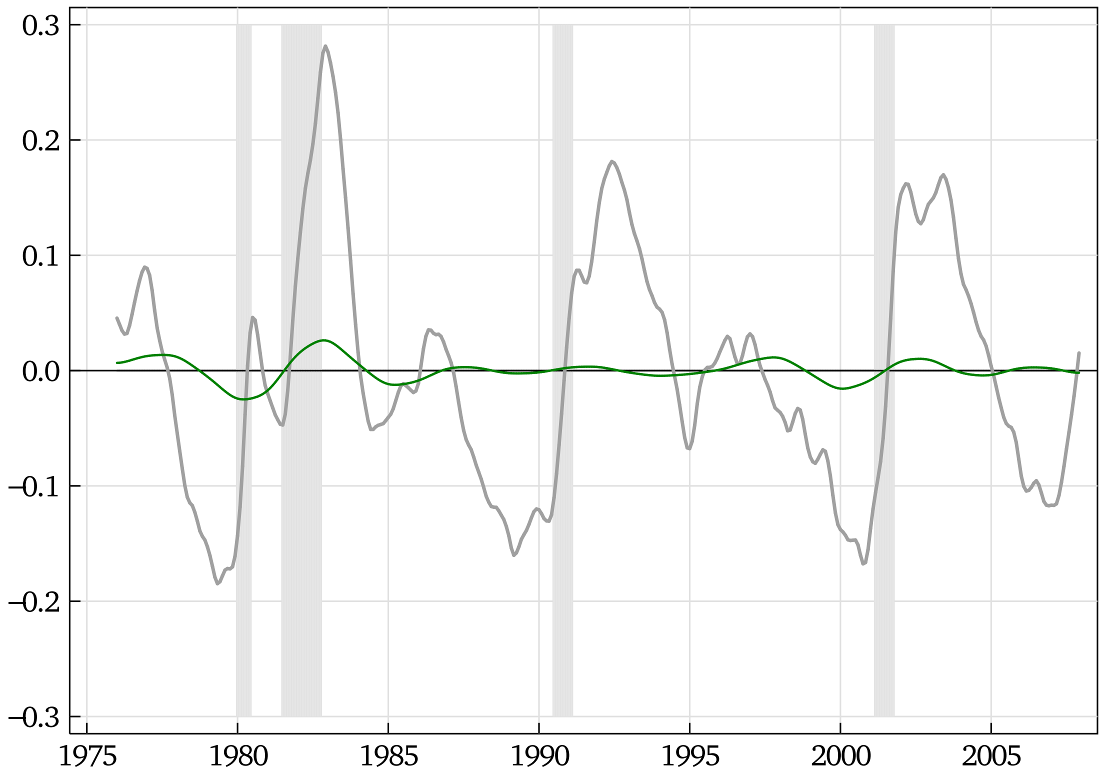

Figure 3.4. Cyclical Behavior of Geographic Mobility, 1976–2007[a]

Source: Author's calculations using annual data from U.S. Census Bureau (2007)

a. Cyclical component estimated using equation 3.8. Gray line is cyclical component of unemployment rate. Annual data are smoothed to monthly frequency using locally weighted polynomial regression smoother (Cleveland, 1979). Shaded regions indicate recessions as dated by the National Bureau of Economic Research (NBER).

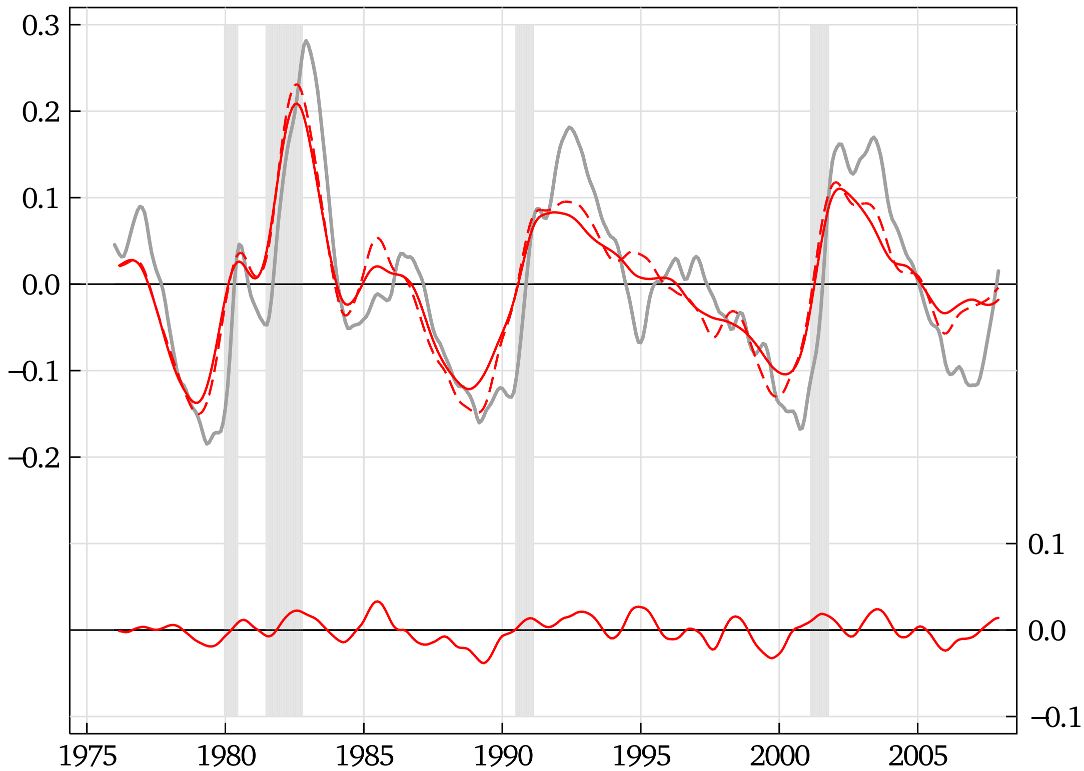

Figure 3.5. Effect of Mobility on Cyclical Component of Separation Hazard Rate, 1976–2007[a]

Source: Author's calculations using CPS microdata for 1976:2–2007:12.

a. Cyclical component estimated using equation 3.8. Gray line is cyclical component of unemployment rate. Thin solid line uses entire population; dashed line uses only population of stayers. Lower panel plots difference between cyclical component estimated from population of stayers and that of entire population; see equation 3.9. Shaded regions indicate recessions as dated by the NBER.

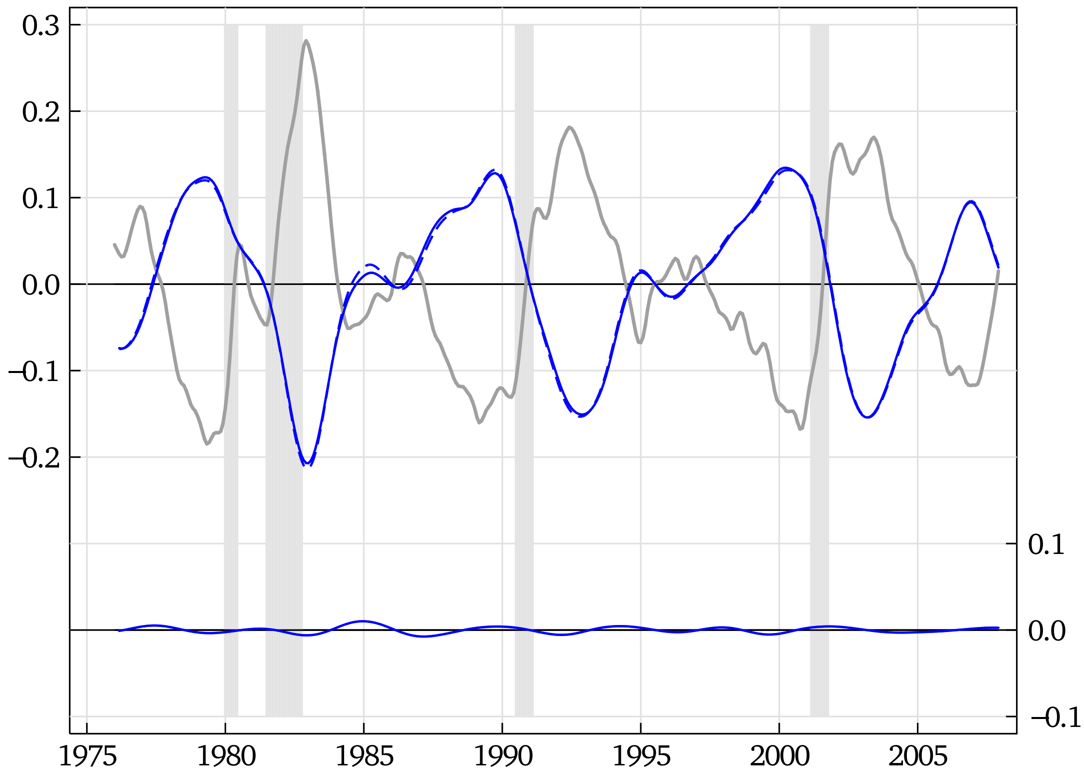

Figure 3.6. Effect of Mobility on Cyclical Component of Job Finding Hazard Rate, 1976–2007[a]

Source: Author's calculations using CPS microdata for 1976:2–2007:12.

a. Cyclical component estimated using equation 3.8. Gray line is cyclical component of unemployment rate. Thin solid line uses entire population; dashed line uses only population of stayers. Lower panel plots difference between cyclical component estimated from population of stayers and that of entire population; see equation 3.9. Shaded regions indicate recessions as dated by the NBER.

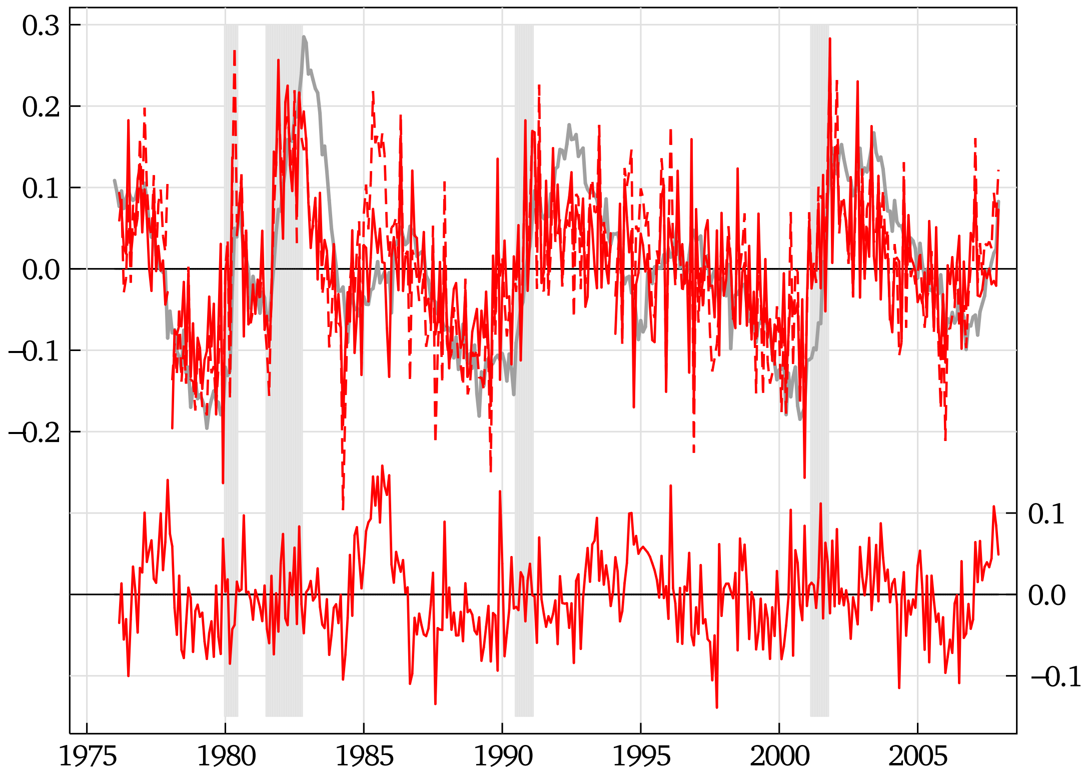

Figure 3.7. Effect of Mobility on Cyclical Component of Separation Hazard Rate, Alternate Measure, 1976–2007[a]

Source: Author's calculations using CPS microdata for 1976:2–2007:12.

a. Cyclical component isolated from seasonally-adjusted series using Hodrick-Prescott filter. Gray line is cyclical component of unemployment rate. Solid maroon line uses entire population; dashed maroon line uses only population of stayers. Lower panel plots difference between cyclical component estimated from population of stayers and that of entire population; see equation 3.9. Shaded regions indicate recessions as dated by the NBER.

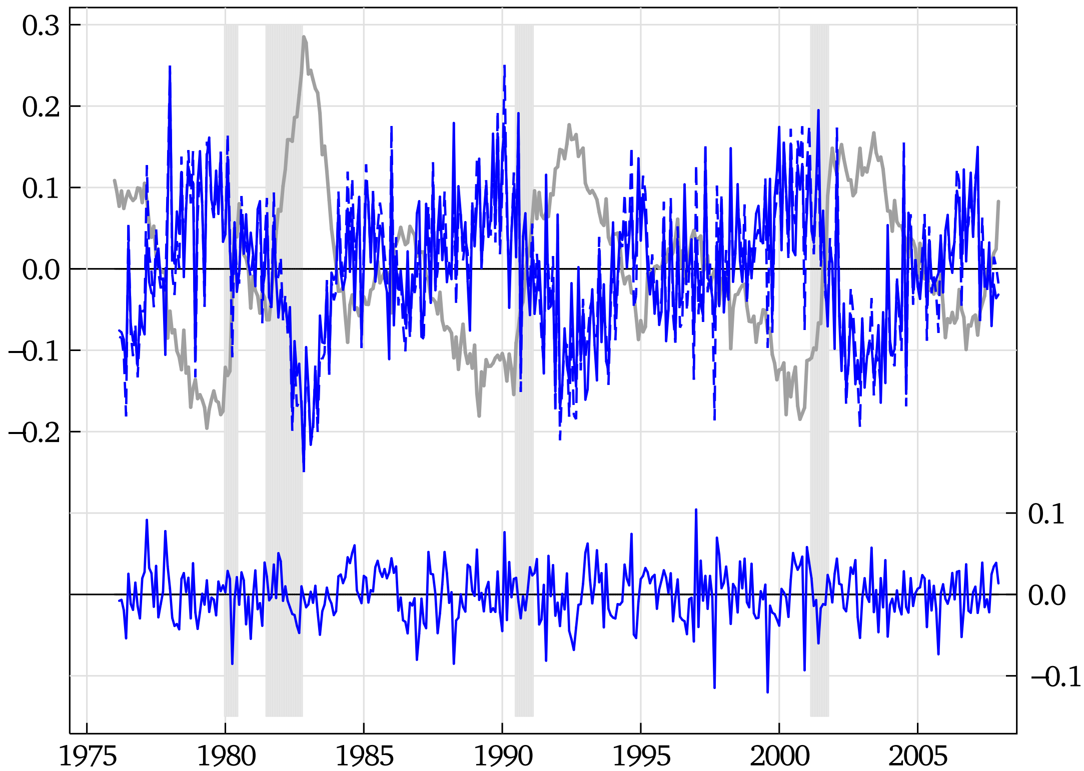

Figure 3.8. Effect of Mobility on Cyclical Component of Job Finding Hazard Rate, Alternate Measure, 1976–2007[a]

Source: Author's calculations using CPS microdata for 1976:2–2007:12.

a. Cyclical component isolated from seasonally-adjusted series using Hodrick-Prescott filter. Dark gray line is cyclical component of unemployment rate. Solid navy line uses entire population; dashed navy line uses only population of stayers. Lower panel plots difference between cyclical component estimated from population of stayers and that of entire population; see equation 3.9. Shaded regions indicate recessions as dated by the NBER.

Chapter 4

Weekly Time Series of the U.S. Labor Market

4.1 Introduction

At least since Kaitz (1970) and Perry (1972), many models of the labor market have taken the week as the fundamental unit of time. Recently, there has been increasing interest in using the weekly frequency for discrete-time search and matching models.[1] Previously, information needed to calibrate a matching model has only been available based on monthly estimates of labor market behavior.

This chapter uses data from the Survey of Income and Program Participation (SIPP) over 1983–2006 to create a new data set of U.S. labor market behavior at weekly frequency. Unlike the Current Population Survey (CPS), which only collects data about a specific week of the month, the SIPP collects information for every week. By applying CPS labor force definitions to these data I can construct weekly time series of the U.S. labor market, including the number of direct employment-to-employment transitions.

Because the SIPP is not designed for aggregate time series analysis, several obstacles must be overcome to create the weekly series. In particular, the SIPP data suffer from a phenomenon known as the "seam effect," whereby transitions

1. See Hagedorn and Manovskii (2008); Ramey (2008); Elsby et al. (forthcoming).

tend to be concentrated at the seam between two waves of interviewing. I devise a correction for the seam effect that allows for consistent estimates of aggregate series. The chapter documents these difficulties and provides a detailed description of how the weekly data are constructed from the SIPP microdata.

I then assess how the labor force measures constructed from the SIPP compare with a known benchmark, the CPS. However, because the weekly SIPP data are not strictly comparable to the CPS, I construct "synthetic" CPS measures within the SIPP that replicate, to the best extent possible, how a SIPP respondent would be classified if surveyed by the CPS. I then compare the SIPP with the CPS along three dimensions: labor force stocks, gross flows, and cyclical dynamics.

The labor force stocks estimated from the SIPP and CPS are very similar in level and are highly correlated. However, the number of transitions among labor force states measured in the SIPP is substantially lower than in the CPS.[2] SIPP gross flows are between one-third and one-half as large as those estimated from the CPS. However, the volatility of gross flows is similar to that in the CPS and the time-series correlation between series from the two data sources is high.

The cyclical dynamics captured by the SIPP are quite similar to those in the CPS. The estimated cyclical components of the separation and job finding hazard rates in the SIPP and CPS have similar time-series behavior and correlation coefficients of 0.6 or higher. The SIPP exhibits significantly larger cyclical volatility. In both data sources the separation hazard rate is strongly countercyclical and the job finding hazard rate is strongly procyclical, although the relationship is weaker in the SIPP.

Thus, at a monthly frequency the SIPP and CPS have similar cyclical dynamics. An advantage of the SIPP over the CPS is that it provides more detailed information about labor market dynamics. In particular, the SIPP data can be used to construct weekly time series of the U.S. labor market. There is concern among researchers that measuring transitions using monthly data may lead to bias from time aggregation.[3] In chapter 2 I show that time aggregation does not lead to cyclical bias in gross flows or hazard rates. However, gross flows estimated from monthly

2. This is similar to findings by Nagypál (2004) and Bils et al. (2007).
3. Shimer (2007).

data understate the true number of transitions by approximately 20 percent when all weekly transitions are measured.

Another benefit is that the SIPP includes information on job changes by workers who remain employed. Fallick and Fleischman (2004) estimate that roughly twice as many workers separate direct to another job without an intervening spell of unemployment. These employment-to-employment (EE) separations do not directly affect unemployment. More recently, Moscarini and Thomsson (2008) argue that the EE rate is even higher.

However, the CPS may overstate direct EE transitions due to time aggregation. In the CPS since 1994, when a person is determined to be employed during the reference week they are asked whether they are employed with the same employer. If they answer no, this is identified as a direct EE transition. However, the CPS does not contain information about her labor market behavior outside the previous week. It may, however, be the case that a person was employed in consecutive reference weeks and with a new employer the second month but did *not* make a direct EE transition.

Using the SIPP allows me to identify direct EE transitions at the weekly level, eliminating time aggregation. Abstracting from labor force participation, I construct new measures of the EE and EU transition rates at weekly frequency. I find that employment-to-employment transitions account for one-half of all separations from employment. This estimate is 50 to 60 percent smaller than estimates of direct EE transitions using the CPS.[4]

Finally, I examine the cyclical behavior of the labor market at weekly frequency. The EE and EU rates have roughly the same volatility as unemployment at business cycle frequencies. In contrast, the total separation rate, the sum of the EE and EU rates, is substantially less variable than either of its components and only 60 percent as volatile as output. The EU rate is strongly countercyclical, leading unemployment by ten months, while the EE rate has a strong negative correlation and leads unemployment by five months. The combination yields a nearly acyclical total separation rate. Thus, the apparently weak cyclical movements of the total separa-

4. Fallick and Fleischman (2004); Moscarini and Thomsson (2008).

tion rate mask strong movements in underlying separation activity at the EE and EU margins. The weekly job finding rate is almost twice as volatile as unemployment over the business cycle. It is strongly procyclical and coincident with unemployment.

The chapter proceeds as follows. Section 4.2 discusses the SIPP and the technical aspects of how the weekly data set is constructed. In particular, section 4.2.5 describes the seam effect and the correction I devise. In section 4.3 I compare the labor force measures constructed from the SIPP with the CPS. Section 4.4 analyzes weekly hazard rates and the cyclical behavior of total separations at weekly frequency. The final section concludes.

4.2 Survey of Income and Program Participation

The Survey of Income and Program Participation (SIPP) is an ongoing longitudinal survey of U.S. households. It is similar in many respects to the CPS, allowing for concurrent analysis, yet the SIPP offers researchers additional information and richness not available in the CPS. This section describes the SIPP survey, focusing on elements important for estimating aggregate time series.

4.2.1 Survey Design

The largest organizational unit of the SIPP is the *panel*. Each SIPP panel is formed from a nationally-representative sample of individuals fifteen years of age and older selected from households in the civilian noninstitutional population. These sampled individuals, along with others who subsequently live with them, are interviewed once every four months over the duration of the panel. Unlike the CPS, original sample members fifteen years or older who move from their original address to another address are interviewed at the new address. If persons not previously in the survey join a respondent's household, they are interviewed for as long as they live with the original respondent.

Each panel is randomly divided into 4 *rotation groups*, with each rotation group interviewed in a separate month. For a given panel, a set of interviews conducted for each of the 4 rotation groups constitutes 1 interview *wave*. At each interview respondents are asked to provide information about the previous four months. Table 4.1 shows the relationship between rotation group, calendar month, and survey wave for the 1985 panel. For example, the first column of table 4.1 shows that rotation group 1 had its wave 1 interview in January 1985. At that interview respondents provided information about the previous four months, beginning with October 1984.

Each rotation group is interviewed the same number of times in each panel. However because each rotation group enters the survey universe in a different month, each rotation group spans a different set of calendar months. This can be seen clearly in table 4.1. Note that the first and last three months of every panel do not have observations from all 4 rotation groups; this becomes important when considering the seam effect (discussed below).

Each panel was originally designed to have 8 waves of interviews and a target initial sample size of 20,000 households. In practice, however, insufficient funding led to the early termination of several panels and frequent shortfalls in the target size. The original SIPP survey design also called for a new panel to begin each year, giving an overlapping design to improve accuracy.

The SIPP survey underwent a substantial redesign in 1996 to improve the quality of longitudinal estimates. The overlapping panel structure was eliminated in favor of a substantially larger sample size and panel length was increased from thirty-two months to forty-eight months. In addition, computer-assisted survey techniques, such as dependent interviewing, were introduced.[5]

Currently there are data available for 13 SIPP panels. The 1989 panel, which only lasted 3 waves, is not used. The time series coverage of each panel is shown in figure 4.1. Each of the 12 SIPP panels used in this chapter is shown on the vertical axis. Each month a panel contributes data is indicated by a solid line.

The first panel is the 1984 panel, although data from the first interview go back to June 1983. A new panel is added each year until 1993. The 1996 panel is the first selected under the new survey design; data begin in December 1995. There is a seven month period from March 2000 to September 2000 where no interviews

5. Similar improvements were implemented in the CPS following its redesign in 1994. See Bureau of Labor Statistics (2002).

were conducted because of insufficient funding. Data from the 2001 panel begin in October 2000.

4.2.2 Survey Content

The core content of the survey consists of questions asked at every interview, covering demographic characteristics; labor force participation; program participation; amounts and types of earned and unearned income received, including transfer payments; noncash benefits from various programs; asset ownership; and private health insurance.[6] Most core data are measured on a monthly basis. Some core items are recorded only once per wave (e.g. race), while others are measured on a weekly basis (e.g. labor force status).

The information necessary to calculate gross flows is contained in 2 types of microdata files: full panel files and core wave files. Core wave files are released following the completion of a survey wave and contain the core labor force data and individual sampling weights. Wave files generally contain one record for each person in each month of the wave (e.g. up to four records per wave for each sample member).

Full panel files are released after interviewing for an entire panel is completed. They contain one record for each person interviewed at any time during the panel.

The full panel files are the best choice for longitudinal analysis. They contain demographic information for each person in the sample that has been edited to ensure longitudinal consistency. Missing observations from persons who were not interviewed for 1 or more months are either imputed or are identified as not in the sample.

Unfortunately, the full panel files have two major drawbacks for constructing gross labor flows. First, individuals' records are indexed by reference month, not calendar date of the interview. Because each rotation group begins in a different month there is not a one-to-one correspondence between reference month and calendar month within each panel file. Second, the full panel files do not contain

6. Westat (2001), p. 1-4.

sufficiently detailed information on labor force classification and sampling weight. These issues are addressed in section 3.

4.2.3 SIPP Data Sources

As discussed above, the SIPP data come in two forms: full panel files and core wave data. The full panel files contain edited and longitudinally-consistent demographic information. They form the basis for defining a person's observations in the SIPP. The necessary labor force and sample weight information from the core wave files is then merged into the full-panel file.

Individuals are matched longitudinally using 3 variables: the sample unit identifier (`SSUID`), the entry address identifier (`EENTAID`), and the person number (`EPPPNUM`).[7]

There are three groups of SIPP panels over which the data structures and procedures are consistent: 1984–1988, 1990–1996, 2001. I discuss each period in turn.

Panels 1984–1988

The Census Bureau publishes full panel files for each of the 5 panels in this period (1984, 1985, 1986, 1987, 1988). A full panel file was not produced for the dramatically-shortened 1989 panel; few usable observations are lost by excluding it. The following variables are taken from the full panel files: the three identification variables, rotation group, interview status, sex, and age. The ID, rotation group, and sex variables are constant across the panel for each person but interview status and age can change in each month. Note that no information about the calendar date is contained in the full panel file.

The core wave files for this period have a "rectangular" structure (i.e. 1 observation per person per wave) and must be reshaped to a person-month format (i.e. 4 observations per person per wave). For each person, the following variables are taken from the core wave files: the three identification variables, date, sampling

7. All variables are named using the 1996 panel definitions.

weight, and 3 labor force variables.

Because the labor force recode is not available on a weekly basis prior to the 1990 panel, it must be constructed using the answers to 3 weekly questions:

1. Did this person have a job or business during this week of the reference period? (`WKWJOB`)

2. Was this person with a job or business but without pay for this week of the reference period? (`WKWABS`)

3. Was this person looking for work or on layoff during this week of the reference period? (`WKLOOK`)

A weekly labor force recode consistent with CPS definitions is constructed by the following rules:

1. A person is *employed* if $\texttt{WKWJOB} = 1$ or if $\texttt{WKWJOB} = 0$ and $\texttt{WKWABS} = 1$;

2. A person is *unemployed* if $\texttt{WKWJOB} = 0$ and $\texttt{WKLOOK} = 1$; and,

3. A person is *not in the labor force* if $\texttt{WKWJOB} = 0$ and $\texttt{WKLOOK} = 0$.

The constructed labor force recode variable and all other variables from the core wave files are then merged into the full panel file to create the dated time series for each person.

Panels 1990–1993

The full panel files are available for the 1990, 1991, 1992, and 1993 panels. The same 8 variables are extracted from the full panel files as for the previous period. The core wave files for this period are published in person-month format and require no reshaping. The same 5 non-labor force variables are taken from the core wave files. A change in the weekly labor force coding allows for direct extraction of the weekly labor force recode.

The weekly labor force recode for week w ($\texttt{WKESR}w$) classifies persons into 5 states. The CPS-equivalent labor force status is given by:

1. A person is *employed* if WKESR$w = 1, 2$, or 3;
2. A person is *unemployed* if WKESR$w = 4$; and,
3. A person is *not in the labor force* if WKESR$w = 5$.

The constructed labor force recode variable and all other variables from the core wave files are then merged into the full panel file to create the dated time series for each person.

Panels 1996, 2001, and 2004

No full panel files are published for the panels after 1993. For 1996, "panel longitudinal" core wave files, which have undergone longitudinal editing similar to full panel files, are published. Only core wave files are available for the 2001 and 2004 panels. All variables are taken from these core wave files. The labor force classification follows that for the previous period.

4.2.4 Constructing Aggregate Time Series

When estimating a longitudinal object such as gross flows, each rotation group should be thought of as its own separate panel—where here "panel" has its traditional econometric meaning: a collection of repeated observations on the same cross-section of individuals. Because each SIPP panel is nationally representative and because households are randomly assigned to rotation groups, the SIPP data can be viewed as 48 smaller, overlapping panels.

Let $p = 1, 2, \ldots, 12$ index SIPP panels and $r \in \{1, 2, 3, 4\}$ index the rotation group within a SIPP panel. An individual rotation group is uniquely identified by pr. In month t there are observations from P_t panels, each with R_{pt} rotation groups. Let $j = 1, 2, \ldots, m_{prt}$ index persons from rotation group pr in month t.

The estimator of a population total for some data object Y from rotation group pr for month t is given by

$$\widehat{Y}_{prt} = \sum_{j=1}^{m_{prt}} w_{prjt}\, y_{prjt}\,, \tag{4.1}$$

where y_{prjt} is the individual's response for object Y and w_{prjt} is his sampling weight for month t.[8] The population estimator for rotation group pr is the weighted sum of responses for all persons in that rotation group. The aggregate estimate for month t is taken across all panels and rotation groups:

$$\widehat{Y}_t = \sum_{p=1}^{P_t} \sum_{r=1}^{R_{pt}} \sum_{j=1}^{m_{prt}} \omega_{prt} w_{prjt}\, y_{prjt}\,. \tag{4.2}$$

Each rotation group is weighted by its contribution to the total number of observations in a month:

$$\omega_{prt} = \frac{N_{prt}}{\sum_{p=1}^{P_t} \sum_{r=1}^{R_{pt}} N_{prt}}. \tag{4.3}$$

The pooled estimates are found by further aggregating over time:

$$\widehat{Y} = \sum_{t=1}^{T} \widehat{Y}_t. \tag{4.4}$$

Equations 4.1 and 4.4 are estimated separately for each of the 3 labor force stocks and 9 transitions. For stocks, the object Y is an indicator for having labor force status `I` in month t. For example, when estimating the stock of employed $y_{prjt} = I(\texttt{LFS}_{prjt} = \texttt{E})$. For labor force transitions, the aggregation of all individual `ij` transitions is called the `IJ` *flow*, where capital letters indicate the aggregate quantity. Thus `IJ` is the number of persons who move from state I in month $t-1$ to state J in month t.

4.2.5 The Seam Effect

A phenomenon known as the "seam effect" is a well-documented but little-understood problem in the SIPP.[9] A *seam* in the SIPP is the boundary between four-month reference periods in successive waves of a panel.[10] The seam effect is characterized by observing significantly more changes in survey variables from

8. Individual sampling weights reflect a person's weight in the entire SIPP panel. These are multiplied by 4 to make each rotation group nationally representative.
9. See Westat (1998), chapter 6.
10. Westat (1998), p. 63.

the last month of the previous wave to the first month of the next wave (i. e., at the seam) than between any two months within a wave. The seam effect is prevalent in measures of labor force behavior, particularly in identifying a change of employer.

Exactly why the SIPP suffers from seam effects have not been definitively identified. Westat (1998) finds that research on the seam phenomenon in recipiency items has no association with the characteristics of respondents, edits and imputations, proxy versus self-response, or changes in interviewer assignments.[11] This suggests that the seam effect may be the product of inertia in reporting or recall bias.

Figures 4.2 and 4.3 illustrate the seam effect in two labor force transition measures estimated from the 1996 panel. Each panel graphs the flow estimated separately for each rotation group (thin line) together with the average across all rotation groups (thicker line) that excludes rotation groups on a seam. All panels share common time- and y-axes to facilitate comparison. Dashed vertical lines indicate the seam between waves.

Figure 4.2 plots the gross flows from employment to unemployment. There are obvious and sizable jumps in separations to unemployment estimated at the seam compared with non-seam months, although the seam effect is somewhat obscured by seasonal variation. Comparing the separation flow estimated from rotation group 3 to the average across all rotation groups illustrates the large deviations occurring at seams.

The seam effect is particularly severe for employment-to-employment transitions (figure 4.3). The observation on the seam records 2 to 4 times as many flows as the nonseam observations. Job-to-job transitions occur when a person reports working for a different employer without a change in labor force status. The survey instrument asks respondents for the date of the change in employer ID, so it is surprising that dramatically more employment-to-employment changes are reported at the seam. This indicates that the SIPP measure of direct employment-to-employment (EE) flows will be sensitive to correcting for the seam effect.

Although the seam effect is typically described as a monthly phenomenon,

11. Westat (1998), p. 64.

it manifests itself at the first transition between waves, regardless of frequency. For measures calculated at monthly frequency, the seam occurs at the first month of the wave. Thus, because of the SIPP's rotation pattern, only one-fourth of the sample is on a seam for any pair of calendar months within a rotation group. For the weekly SIPP data, the seam phenomenon affects only the first week of each wave.

I test for the presence of the seam effect using a fixed-effects regression on data aggregated at the rotation group–level. Let $s_{prt} = 1$ if rotation group pr is on a seam at week t and $s_{prt} = 0$ otherwise. To test for the seam effect in object Y, I regress $\ln(Y_{prt})$ on s_{prt} and fixed effects for panel, rotation group, and time:

$$\ln\left(Y_{prt}\right) = \alpha_0 + \alpha_{1p} I(p) + \alpha_{2r} I(r) + \alpha_{3m} I(m) + \beta s_{prt} + \xi_{prt}\,, \tag{4.5}$$

where Y_{prt} is the stock- or flow-population ratio and where $I(p)$, $I(r)$, and $I(m)$ represent fixed effects for panel, rotation group, and month. The omitted groups are $p = 1984$, $r = 1$, and $m =$ June 1986. Table 4.3 reports the estimated coefficient $\widehat{\beta}$ for the SIPP labor force stocks and gross flows.

Looking first at stocks (upper panel), the presence of a seam effect is rejected in all 3 stocks. As suggested by Westat (1998), response variance on the seam does not affect cross-sectional estimates such as labor force stocks. The seam affects the *composition* of the transitions that cumulate to the period t stock.

The lower panel of table 4.3 confirms the visual evidence of the large seam effects from the earlier figures. All coefficient are positive, indicating more transitions recorded on a seam, and highly statistically significant. Direct EE flows are 3.6 times higher on a seam than in nonseam weeks. The same pattern holds for the other flows, with coefficients indicating 2–3 times greater flows on a seam.

Although the seam effect occurs regularly, it still is necessary to correct for it in aggregate estimates because of periods when no seams occur. This occurs in the first 3 months of each panel and at any time a rotation group is missing from the sample. During the months when no seam is present, such as a junction between two panels, a dramatic decline in measured flows is observed when using uncorrected data. There are four periods in the SIPP where a non-overlapping junction between panels produces no seams: in 1990, 1996, 2000, and 2004.

4.2.6 Seam Effect Correction

The previous sections shows large and statistically significant seam effects in all flows. To mitigate the effect on aggregate estimates, I correct the data for the seam effect.

Because the SIPP's rotation pattern places only one-fourth of the sample on a seam for any month, it is possible to infer the true behavior using the rotation groups not on a seam. Consider an alternate estimate of object Y that is calculated excluding any rotation groups on a seam:

$$Y_t^{ns} = \sum_{p=1}^{P_t} \sum_{r=1}^{R_{pt}} \omega_{prt}^{ns} (1 - s_{prt}) Y_{prt} \, , \tag{4.6}$$

where $s_{prt} = 1$ if rotation group pr is on a seam at date t and equals zero otherwise and weight ω_{prt}^{ns} is defined analogously to equation 4.3 for nonseam observations. In a sufficiently large sample, equations 4.6 and 4.1 are both consistent estimators for Y_t.

I use equation 4.6 to calculate aggregate estimates, in effect replacing observations at a seam with the average of all nonseam observations for that date. Because seams occur at fixed intervals and are uncorrelated with the sample selection, this missing-at-random (MAR) correction is consistent.

4.3 Comparing SIPP and CPS Labor Force Measures

This section assess how the labor force measures constructed from the SIPP compare with a known benchmark, the CPS. I first compare the SIPP with the CPS along three dimensions: labor force stocks, gross flows, and cyclical dynamics. However, because the weekly SIPP data are not strictly comparable to the CPS, I first discuss constructing "synthetic" CPS measures within the SIPP; these measures replicate to the best extent possible, how a SIPP respondent would be classified if surveyed by the CPS. CPS data are described in chapter 3.

4.3.1 Synthetic CPS Labor Force Classification

The CPS determines an individual's labor force status for a month based on his experience during that month's *reference week*. The SIPP monthly labor force recode is not strictly comparable to the CPS measure. However, it is possible to construct a "synthetic" CPS labor force classification using weekly labor force classification described in section 4.2.3.[12] This labor force classification is the closest possible measure to what the person would have been classified were he surveyed by the CPS.

The SIPP core wave files contain questions pertaining to labor force status for each week of the reference period. The first step in constructing the synthetic CPS labor force classification is to identify the CPS reference week within each wave file.

The CPS reference week is defined as the 7-day period, Sunday through Saturday, that includes the 12th of the month. In December, the week of the 5th is used as the reference week, provided that the week falls entirely within the month; otherwise the week containing the 12th is used as the reference week.

After identifying the CPS reference week for each month the weekly SIPP labor force information can be used to determine the individual's CPS labor force classification. The correspondence between the SIPP weekly labor force recode (`WKESR`)—or the constructed weekly labor force recode—and the CPS labor force definitions is provided in the previous section. A person's CPS labor force classification is defined as the corresponding labor force recode in the CPS reference week.

The synthetic CPS recode is harder to construct for the pre-1990 panels. The core wave files before 1990 organize the weekly information chronologically by week (i.e. `WKESR1`, `WKESR2`, `WKESR4`, ..., `WKESR18`). However because each rotation group begins in a different month, `WKESR1` for rotation group 1 does not represent the same calendar week as `WKESR1` for rotation group 2. Thus a correspondence between SIPP reference week and calendar week must be determined separately for each rotation group.[13]

12. This is the same procedure developed in Fujita et al. (2007).

13. Identifying this correspondence is further complicated because several waves have only 3 rotation groups and the 1984 panel wave 8 file does not contain data for rotation group 3.

This is best illustrated with an example. In wave 1 of the 1986 panel, the first observation for rotation group 2 is for January 1986. The CPS reference week for this month was week three, so the synthetic CPS labor force status for January 1986 is determined by `WKESR3`. The reference week for February was the second week and the corresponding CPS reference week is `WKESR7` (January 1986 has five weeks plus the two weeks until the reference week). The remaining reference weeks are calculated similarly.

Classification as employed and not in the labor force (NILF) follow directly from the CPS definitions; unemployment requires an additional step. The CPS classifies a person as unemployed if he has searched for a job within the last four weeks. For this section only I apply this definition on a rolling basis to determine a person's weekly synthetic CPS labor force status. That is, a person without a job would be considered unemployed this week if he had searched for work during any of the previous four weeks, even if he did not search this week. After four weeks without search has elapsed, a person is classified as NILF. For the weekly labor force measures used in section 4.4, a person is only considered unemployed in a week where he is searching for a job.

Labor force transitions are measured by comparing a person's labor force status in two successive time periods. I define a transition from state `i` in period $t-1$ to state `j` in period t as an `ij` transition observed at t. Transitions are identified at two different frequencies. A person's *weekly* labor force transition is the change in labor force status from one week to the next week. A person's *monthly* labor force transition is the change in labor force status from one CPS reference week to the next CPS reference week. This "synthetic" CPS labor force transition records how a person would have been classified by the CPS.

4.3.2 Stocks

The starting point for comparison is the estimates of the civilian noninstitutional population aged sixteen and older derived from both data sources. The time averages of the estimated population in the SIPP and the CPS are virtually identical and the correlation between the two population levels is 0.9966. For the rest of the analysis, I compare population ratios for the objects of interest rather than absolute levels.

Table 4.4 reports the labor force stocks, expressed as a share of population, for the CPS and the SIPP. All statistics are calculated for the SIPP sample period (June 1983–December 2006) from seasonally unadjusted data. The sample averages for all three stocks are very similar. The two data sets agree on employment, with nearly identical averages and similar volatility. The SIPP data have about 7 percent fewer unemployed and correspondingly more persons NILF. In addition, the SIPP unemployed stock is over 40 percent more volatile than the CPS stock. However, the correlation between the SIPP and CPS stocks is highest for unemployment (0.95). Employment is similarly strongly correlated in the two data sets (0.91) although the correlation for NILF is weaker (0.73).

Figure 4.4 plots the stocks estimated from the SIPP and CPS. The figure shows CPS data for 1976–2007 (dashed line) together with SIPP data (solid line). The upper panel confirms the similarity of employment in the SIPP and CPS. The unemployment series (middle panel) also both track each other closely. The SIPP reports more unemployed than the CPS over 1983–1990 and slightly fewer during 1996–2001. Both show a similar increase during periods surrounding the 1991 and 2001 recessions.

The stocks of persons NILF do not agree as well as those for employment and unemployment. In particular, the SIPP series is too high in the 1990–1993 panels, although it returns to the CPS level with the 1996 panel. There is also a significant disagreement in the `N` stock at the start of the 2004 panel; there is a corresponding jump in employment.

4.3.3 Gross Flows

Table 4.5 compares the gross flows measured from SIPP and CPS. The first 2 columns report the mean and standard deviation of the CPS gross flows, expressed as a percent of the population. Labor force transitions are relatively rare events, accounting for only about 7 percent of all observations; the remaining 93 percent of observations record no change in labor force state.

The next 3 columns report data for the SIPP gross flows. For comparison, the level and volatility are expressed relative to the CPS values. As discussed in section 4.2.5, it is important to correct the SIPP gross flows for seam effects. The gross flows reported in table 4.5 using the seam effect correction from section 4.2.6.

As has been noted elsewhere, the SIPP records fewer labor force transitions than the CPS.[14] The SIPP gross flows from employment to unemployment are only 43 percent as large as those from the CPS while flows from unemployment to employment are just over one-half as large. Although the SIPP levels are substantially lower, the volatility is similar to the CPS and the correlation of the series in the two data sources is much higher (columns 4 and 5 of table 4.5).

Despite the SIPP having a larger measured N stock, the gross flows between employment and nonparticipation are lower relative to the CPS than for unemployment. Flows to and from nonparticipation are roughly one-third as large in the SIPP as in the CPS. As with transitions involving unemployment, the relative volatility is similar and the gross flows tend to move together in two data sets.

Although the number of transitions measured in the SIPP is only one-third to one-fifth as large as in the CPS, the time-series behavior of the series are similar. Although the time-series correlation of EU and UE flows between the SIPP and CPS is high, the implications for the cyclical dynamics must be assessed using a more sophisticated method.

4.3.4 Cyclical Dynamics

To compare the cyclical dynamics of the SIPP and CPS, I focus on the dynamics of the separation and job finding hazard rates. The monthly separation and job finding hazard rates are calculated by

$$s_t = \frac{\mathtt{EU}_t}{\mathtt{E}_{t-1}} \quad \text{and} \quad f_t = \frac{\mathtt{UE}_t}{\mathtt{U}_{t-1}} \tag{4.7}$$

where E and U are the stock of employed and unemployed persons.

The mean of the separation and job finding hazard rates in the CPS, expressed in percent, is reported in column 1 of table 4.6. Consistent with the lower

14. Nagypál (2004); Bils et al. (2007); Fujita et al. (2007); Moscarini and Thomsson (2008).

level of gross flows, the mean of the SIPP hazard rates (column 3) is lower than in the CPS. The SIPP separation rate is 55 percent of that in the CPS while the mean job finding hazard rate is 70 percent as large as in the CPS.

Again, although the levels are quite different, the principal concern is how the *cyclical* behavior of hazard rates compares. To assess the cyclical dynamics, I first must isolate the component of the time series that moves at business cycle frequencies. I model the observed time series as the sum of four independent, unobserved components: a trend, a cycle, a seasonal, and an irregular component.[15] The trend represents low-frequency movements in the series. The cyclical component is a stochastic periodic function of time with a frequency at that of the business cycle. The seasonal component represents fluctuations that repeat annually and the irregular component captures the remaining non-systematic variation.

The unobserved-components model for the natural logarithm of a time series Y_t, denoted y_t, is

$$y_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t\,, \tag{4.8}$$

where μ_t is the trend, ψ_t the cyclical, γ_t the seasonal, and ε_t the irregular component. Details of the econometric specification of the components are provided in appendix B.

Equation 4.8 is recast as a state space model where the unobserved components are represented by the state of the system. The unknown parameters are estimated by maximum likelihood using the Kalman filter to update and smooth the unobserved state. The estimation is performed using the STAMP program written by Koopman et al. (2007). The state space form and the details of the estimation appear in appendix B.

Figure 4.5 plots the estimated cyclical components of the separation and job finding hazard rate. The cyclical component of the civilian unemployment rate published by the Bureau of Labor Statistics (BLS) is shown (thick gray line) as an indicator of the business cycle. The solid line shows the estimated cyclical component of the SIPP hazard rate and the dashed line shows the analogous CPS series.

15. This follows the general method described in Harvey (1989).

Several features are apparent. First, except for two periods, the cyclical components of the two data sources track each other reasonably well. In the upper panel, both separation rates rise with unemployment during both the 1991 and 2001 recessions. The cyclical component of the job finding rate (lower panel) in the SIPP and CPS track very closely until 1996.

Two significant disruptions in the cyclical component of both SIPP series are readily apparent. The first occurs at the junction of the 1993 and 1996 panels (October 1995–February 1996) and the second occurs at the junction of the 2001 and 2004 panels (October 2003–February 2004). In both cases, these periods feature incomplete overlap between the panels, resulting in fewer than 4 rotation groups in any month. Although the effects are more abstruse than a discrete jump in the series, the time series model allows for a level shift in the trend component in January 1996 and January 2003 to help mitigate this effect in the cyclical component.[16]

The estimated cyclical components show a substantial oscillation in 1996 and a smaller one in 2003. If one visually smoothes over those two periods, however, the overall relationship between the cyclical dynamics in the SIPP and the CPS are similar. Largely because of these two periods, the standard deviation of the cyclical component of the SIPP hazard rates are nearly twice those of the CPS hazard rates (table 4.6). Nevertheless, the cyclical components of the CPS and SIPP series still have a high correlation: 0.62 for separation hazard rate and 0.82 for the job finding hazard rate.

The final assessment of the cyclical dynamics of SIPP is the correlation of the separation and job finding hazard rates with the business cycle. I use the civilian unemployment rate as in indicator of the business cycle. I estimate the cyclical component of the unemployment (shown in figure 4.5) rate using equation 4.8.

A richer picture of the cyclical dynamics are revealed by a plot of the cross-correlations between the cyclical component of the hazard rate and the cyclical indicator. This shows not only the contemporaneous correlation but also how time aggregation relates to the business cycle at other horizons. I calculate the cross-

16. Interestingly, the period between the 1996 and 2001 panels (March 2000–October 2000) does not suffer from such a disturbance. The cyclical component for this period is estimated directly from the unobserved state of the model using the Kalman smoother.

correlation between the cyclical component of the unemployment rate in month t and $j = 0, 1, \ldots, 24$ leads and lags of the hazard rate, corr $\left(\widehat{\psi}_t^{UR}, \widehat{\psi}_{t+j}^{hr}\right)$, where $\widehat{\psi}_t^{hr}$ is the cyclical component of the hazard rate.

The cross-correlations of the SIPP and CPS hazard rates are shown in figure 4.6. The cross-correlations confirm the strong visual relationship between the cyclical components of SIPP and CPS hazard rates. The contemporaneous correlation ($j = 0$) of the separation hazard rate with unemployment in the CPS data is 0.89. Despite the two visible disruptions in the SIPP series, the contemporaneous correlation of the SIPP hazard rate with unemployment is 0.55, about 38 percent lower. Both data sources find countercyclical separation hazard rates.

The SIPP and CPS agree more closely on the job finding rate. The contemporaneous correlation in the CPS is -0.93 and about 17 percent lower (-0.78) in the SIPP. Job finding is strongly procyclical in both the SIPP and CPS.

4.3.5 Discussion

The stocks of employed and unemployed estimated from the SIPP and CPS are very similar in level and are highly correlated. The number of transitions measured in the SIPP is substantially lower than in the CPS: SIPP gross are between one-third and one-half as large as those estimated from the CPS. However, the volatility of gross flows is similar to that in the CPS and the time-series correlation between series from the two data sources is high.

The cyclical dynamics captured by the SIPP are quite similar to those in the CPS. Although the SIPP time series are obscured by two periods where incomplete overlap between the panels results in significant instability, this disruption can be minimized using the unobserved-components model. The estimated cyclical components of the separation and job finding hazard rates in the SIPP and CPS have similar time-series behavior. In both data sources the separation hazard rate is strongly countercyclical and the job finding hazard rate is strongly procyclical, though the relationship is weaker in the SIPP.

Although the SIPP is designed for different purposes than the CPS, the labor force statistics calculated from the SIPP match those from the CPS remarkably well.

Some difference between the two data sources is expected due to sampling variation and minor differences in survey design and definitions. Broadly speaking, however, the SIPP and the CPS capture similar dynamics of the U.S. labor market.

4.4 Weekly Time Series

Thus, at monthly frequency the SIPP and CPS have similar cyclical dynamics. An advantage of the SIPP over the CPS is that it provides more detailed information about labor market dynamics. In particular, the SIPP data can be used to construct weekly time series of the U.S. labor market. I use the SIPP to construct weekly hazard rates.

The SIPP allows me to identify direct EE transitions at the weekly level, eliminating time aggregation. Abstracting from labor force participation, I construct new measures of the EE and EU transition rates at weekly frequency. I find that employment-to-employment transitions account for one-half of all separations from employment, about 50 to 60 percent smaller than estimates using the CPS.

4.4.1 Constructing Hazard Rates

For the hazard rate analysis, I restrict attention to the employed and unemployed only. I construct hazard rate series at weekly frequency and normalize by stocks implied by those flows. This construction is necessary to normalize by stocks that are consistent with the population of interest.

I first estimate weekly gross flows from the SIPP and take a monthly average. I then construct stocks of employed and unemployed from the measured transitions. Given the timing convention for measuring gross flows, the period t stock, $\mathtt{J}_t$, is the sum of flows of persons who end period t in state j.

$$\tilde{\mathtt{E}}_t = \mathtt{EE}_t^{same} + \mathtt{EE}_t^{new} + \mathtt{UE}_t \tag{4.9}$$

$$\tilde{\mathtt{U}}_t = \mathtt{UU}_t + \mathtt{EU}_t \tag{4.10}$$

I define the four hazard rates below in relation to these stocks:

$$EER_t = \frac{\mathtt{EE}_t^{new}}{\tilde{\mathtt{E}}_{t-1}} \tag{4.11}$$

$$EUR_t = \frac{\mathtt{EU}_t}{\tilde{\mathtt{E}}_{t-1}} \tag{4.12}$$

$$TSR_t = \frac{\mathtt{EE}_t^{new} + \mathtt{EU}_t}{\tilde{\mathtt{E}}_{t-1}} \tag{4.13}$$

$$JFR_t = \frac{\mathtt{UE}_t}{\tilde{\mathtt{U}}_{t-1}}. \tag{4.14}$$

These data and hazard rate measures are also used by Nekarda and Ramey (2007) to evaluate a discrete-time weekly matching model with on-the-job search and direct employment-to-employment transitions.

4.4.2 Results

The first column of table 4.7 reports the average weekly hazard rate over 1983–2006. The EU and EE separation rates both average 0.16 percent a week, indicating that direct job change accounts for one-half of all separations from employment. This is considerably at odds with previous estimates from the CPS.

To put the weekly figures into more comparable terms, multiply by 52/12 to get 0.70 percent a month. This rate is considerably lower than the monthly rates for either EU (1.57 percent, table 4.5) or EE separations (2.7–3.2 percent, Fallick and Fleischman (2004) and Moscarini and Thomsson (2008)). However, as tables 4.5 and 4.6 indicate, the SIPP undercounts the number of transitions by roughly 50 percent. I adjust for this systematic undercounting by dividing each weekly hazard rate by the relative mean for the hazard rates reported in table 4.6.[17]

After adjusting for the systematic undercounting of transitions in the SIPP, the monthly separation implied by the adjusted weekly data rate is 1.27 percent, quite close to the CPS estimate. The implied monthly rate of direct EE separations is also 1.27 percent, about 50 to 60 percent lower than those of Fallick and Fleischman (2004) and Moscarini and Thomsson (2008). Thus, even after adjusting for the

17. Both the EU and EE rates are adjusted by the separation hazard rate.

level of transitions, the SIPP finds considerably less direct employment-to-employment change than in the CPS, suggesting substantial time aggregation in the CPS measures of direct employment-to-employment separations.

The second column of table 4.7 reports the standard deviation of the cyclical component of the four hazard rates. The EE and EU rates have roughly the same volatility as unemployment at business cycle frequencies. The EE rate is slightly more volatile than employment (1.12) while the EU rate is slightly less volatile (0.96). In contrast, the total separation rate is substantially less variable than either of its components and only 60 percent as volatile as output.

Figure 4.7 plots the cyclical components of the four hazard rates together with the cyclical component of unemployment rate. The two separation rates in the upper panel move sharply opposite each other, except for the period in 1996 associated with the panel disruption. Both series appear equally as volatile as unemployment.

The lower panel of figure 4.7 plots the total separation and job finding hazard rates. The total separation rate is considerably less volatile than either unemployment or the job finding rate, and has little clear association with the unemployment rate. The job finding rate mirrors the unemployment rate and displays considerably higher volatility.

The last column of table 4.7 reports the contemporaneous correlation of each hazard rate with unemployment while figure 4.8 plots the cross-correlation. The EU rate is strongly countercyclical and leads unemployment by ten months, while the EE rate has a strong negative correlation and leads unemployment by five months. The combination yields a nearly acyclical total separation rate. Thus, the apparently weak cyclical movements of the total separation rate mask strong movements in underlying separation activity at the EE and EU margins. The weekly job finding rate is almost twice as volatile as unemployment over the business cycle. It is strongly procyclical and coincident with unemployment.

4.5 Conclusion

This chapter uses data from the SIPP to create a new data set of U.S. labor market behavior, including the number of direct employment-to-employment transitions, at weekly frequency. When the weekly data are analyzed in a manner that mimics the CPS, the SIPP data replicate many of the features of the U.S. labor market observed in the CPS. The labor force stocks estimated from the SIPP closely match the CPS stocks. The gross flows estimated from the SIPP, however, are at least 50 percent lower than those in the CPS. It is likely that this difference is a by-product of correcting the SIPP data for the "seam effect." Although its sources are not well understood, the seam effect describes the tendency for transitions in the SIPP to be concentrated at the seam between two waves of interviews. Because these seams do not occur in all months, they must be removed to construct a consistent time series; doing so necessarily reduces the level of gross flows. However, the time-series correlation between gross flows from the two surveys is high.

In addition, the cyclical dynamics captured by the SIPP are similar to the CPS. The cyclical components of the separation and job finding hazard rates have similar time-series behavior in both surveys and exhibit the same cyclical patterns: a strongly countercyclical separation hazard rate and a strongly procyclical job finding hazard rate. The notable difference in the cyclical dynamics is the SIPP's significantly larger cyclical volatility. A significant share of the higher volatility comes from two periods where the SIPP series have large oscillations in the cyclical component not seen in the CPS series. These periods, in 1996 and 2004, arise from incomplete overlap between the junction of two panels. Although the univariate structural time series model can partially compensate for this disturbance, a more comprehensive treatment requires panel structural time series estimation. This will also allow for a more sophisticated treatment of the seam effect, improving estimated gross flows.

Analyzing weekly the time series, I find that the rate of separations to unemployment (EU) and of direct employment-to-employment separations (EE) each account for one-half of the rate of total separations from employment. The cyclical volatility of the EU and EE rates is comparable to unemployment, however the total separation rate is substantially less variable. The EU rate is strongly countercyclical while the EE rate is strongly procyclical, yielding a nearly acyclical total separation rate. The apparently weak cyclical movement of the total separation rate masks strong movements in underlying separation activity. The weekly job finding rate is strongly procyclical and almost twice as volatile as unemployment over the business cycle.

My estimates, adjusted for the SIPP's systematic undercounting of transitions, imply a monthly EE separation rate that is 50 to 60 percent smaller than estimates from the CPS.[18] This suggests that as many as half of employment-to-employment transitions recorded by the CPS may not, in fact, have been direct. Because the CPS does not contain information about labor market behavior outside the reference week, it is not possible to differentiate true EE transitions from separate EU and UE transitions that are aggregated into a direct EE transition. Thus, the CPS will overstate direct EE transitions. Indeed, chapter 2 shows that most unrecorded EU and UE transitions are classified in the CPS as continuous employment. An implication of these results for policymakers is that, although direct employment-to-employment transitions are important, the traditional channel of cyclical employment adjustment (unemployment) is equally important.

Much of this literature abstracts from labor force participation. Yet about 60 percent of the flows into and out of employment involve nonparticipation and the cyclical dynamics of NILF flows are distinct from those involving unemployment. Two rich avenues for future research are understanding the cyclical dynamics of the labor force participation decision and modeling high-frequency movements into and out of the labor force.

18. Fallick and Fleischman (2004); Moscarini and Thomsson (2008).

Table 4.1. Relationship between Survey Wave, Rotation Group, and Calendar Date[a]

	Rotation group			
Date	*1*	*2*	*3*	*4*
1984m10	1–1			
1984m11	1–2	1–1		
1984m12	1–3	1–2	1–1	
1985m1	1–4	1–3	1–2	1–1
1985m2	2–1	1–4	1–3	1–2
1985m3	2–2	2–1	1–4	1–3
1985m4	2–3	2–2	2–1	1–4
⋮	⋮	⋮	⋮	⋮
1987m1	8–1	7–4	7–3	7–2
1987m2	8–2	8–1	7–4	7–3
1987m3	8–3	8–2	8–1	7–4
1987m4	8–4	8–3	8–2	8–1
1987m5		8–4	8–3	8–2
1987m6			8–4	8–3
1987m7				8–4

Source: Author's calculations.

a. Cell entry gives the interview wave number and month number within each wave.

Table 4.2. The Survey of Income and Program Participation

			Number of			
Panel	*Begin*	*End*	*Months*	*Weeks*	*Persons*	*Observations*[a]
1984	Jun 1983	Apr 1986	35	153	48,934	5,031,872
1985	Oct 1984	Jul 1987	34	148	33,457	3,330,820
1986	Oct 1985	Mar 1988	30	131	27,330	2,681,937
1987	Oct 1986	Apr 1989	31	135	27,401	2,797,571
1988	Oct 1987	Dec 1989	27	117	27,145	2,432,822
1990	Oct 1989	Aug 1992	35	152	52,256	5,747,440
1991	Oct 1990	Jul 1993	35	152	33,473	3,689,173
1992	Oct 1991	Dec 1994	39	170	46,756	5,694,370
1993	Oct 1992	Dec 1995	39	169	46,747	5,669,515
1996	Dec 1995	Feb 2000	51	221	89,013	12,727,920
2001	Oct 2000	Dec 2003	39	170	80,026	8,499,728
2004	Oct 2003	Dec 2006	39	170	100,105	11,022,347
All	Jun 1983	Dec 2006	276	1,200	612,643	69,325,515

Source: Author’s tabulations using SIPP microdata for 1983:6–2006:12.

a. Weekly.

Table 4.3. Testing for Seam Effects in Weekly SIPP Data

Dependent variable	*Seam effect*	*t statistic*	*No. obs.*	R^2
Stocks				
E	0.0008	1.56	6,851	0.8630
	(0.0005)			
U	0.0028	0.69	6,851	0.9246
	(0.0040)			
N	−0.0017	−1.44	6,852	0.5458
	(0.0012)			
Gross flows				
EE	3.5939***	63.98	5,790	0.4563
	(0.0562)			
EU	2.0073***	63.91	6,793	0.4734
	(0.0314)			
EN	2.6444***	82.82	6,820	0.5781
	(0.0319)			
UE	1.8967***	61.83	6,825	0.4710
	(0.0307)			
NE	2.7729***	90.46	6,817	0.6225
	(0.0307)			
UN	2.6005***	89.08	6,793	0.6305
	(0.0292)			
NU	2.5885***	82.11	6,804	0.5875
	(0.0315)			

Source: Author's regressions using weekly SIPP microdata for 1983:6–2006:12.

a. Reports coefficient $\hat{\beta}$ from regression of $\ln(Y_{prt}) = \alpha_0 + \alpha_{1p}I(p) + \alpha_{2r}I(r) + \alpha_{3m}I(m) + \beta s_{prt} + \xi_{prt}$, where s_{prt} is an indicator for panel p, rotation group r being on a seam at week t; αs are fixed effects for panel, rotation group, and month.

Table 4.4. Comparison of SIPP and CPS Labor Force Stocks, 1983–2006[a]

Percent of population

Stock	*Mean*	*Standard deviation*	*Correlation with CPS*
CPS			
E	62.46	1.51	1.0000
U	3.87	0.74	1.0000
N	33.67	0.94	1.0000
SIPP			
E	62.38	1.63	0.9143
U	3.59	1.04	0.9487
N	34.03	0.73	0.7252

Source: Author's calculations using SIPP microdata for 1983:6–2006:12 and CPS data from chapter 3.

a. Statistics and correlations are based on 276 monthly observations.

Table 4.5. Comparison of SIPP and CPS Labor Force Gross Flows, 1983–2006[a]

Percent of population

			SIPP		
	CPS		*Relative*[b]		
Flow	*Mean*	*Standard deviation*	*Mean*	*Standard deviation*	*Correlation with CPS*
Separation					
EU	0.94	0.16	0.4255	0.7500	0.8167
EN	1.80	0.33	0.3111	0.7576	0.8695
Accession					
UE	1.03	0.17	0.5146	0.8824	0.7250
NE	1.61	0.24	0.3416	1.0000	0.7442
Participation					
NU	0.90	0.14	0.2444	0.7143	0.6023
UN	0.84	0.11	0.1786	0.7273	0.4254

Source: Author's calculations using SIPP microdata for 1983:6–2006:12 and CPS data from chapter 3.

a. Statistics and correlations are based on 276 monthly observations.

b. SIPP estimate divided by CPS estimate.

Table 4.6. Comparison of SIPP and CPS Hazard Rates, 1983–2006[a]

				SIPP	
	CPS		Relative[b]		
Flow	*Mean*	*Standard deviation*[c]	*Mean*	*Standard deviation*[c]	*Correlation with CPS*[c]
Separation	1.51	5.58	0.5497	1.9373	0.6166
Job finding	26.97	6.99	0.6963	1.6896	0.8181

Source: Author's calculations using SIPP microdata for 1983:6–2006:12 and CPS data from chapter 3.

a. Statistics and correlations are based on 276 monthly observations.

b. SIPP estimate divided by CPS estimate.

c. Cyclical component estimated using equation 4.8.

Table 4.7. Weekly Transition Rates and Unemployment Rate, 1983–2006[a]

		Cyclical component[b]		
Rate	*Mean*	*Standard deviation*	*Relative standard deviation*[c]	*Correlation with UR*
EER	0.16	0.1043	1.1167	−0.6563
EUR	0.16	0.0900	0.9636	0.5476
TSR	0.33	0.0558	0.5974	−0.1944
JFR	4.29	0.1693	1.8126	−0.7885
UR	5.82	0.0934	1.0000	1.0000

Source: Author's calculations using BLS data and SIPP microdata for 1983:7–2006:11.

a. Monthly average of weekly hazard rates.

b. Cyclical component estimated using equation 4.8.

b. Relative to standard deviation of unemployment rate.

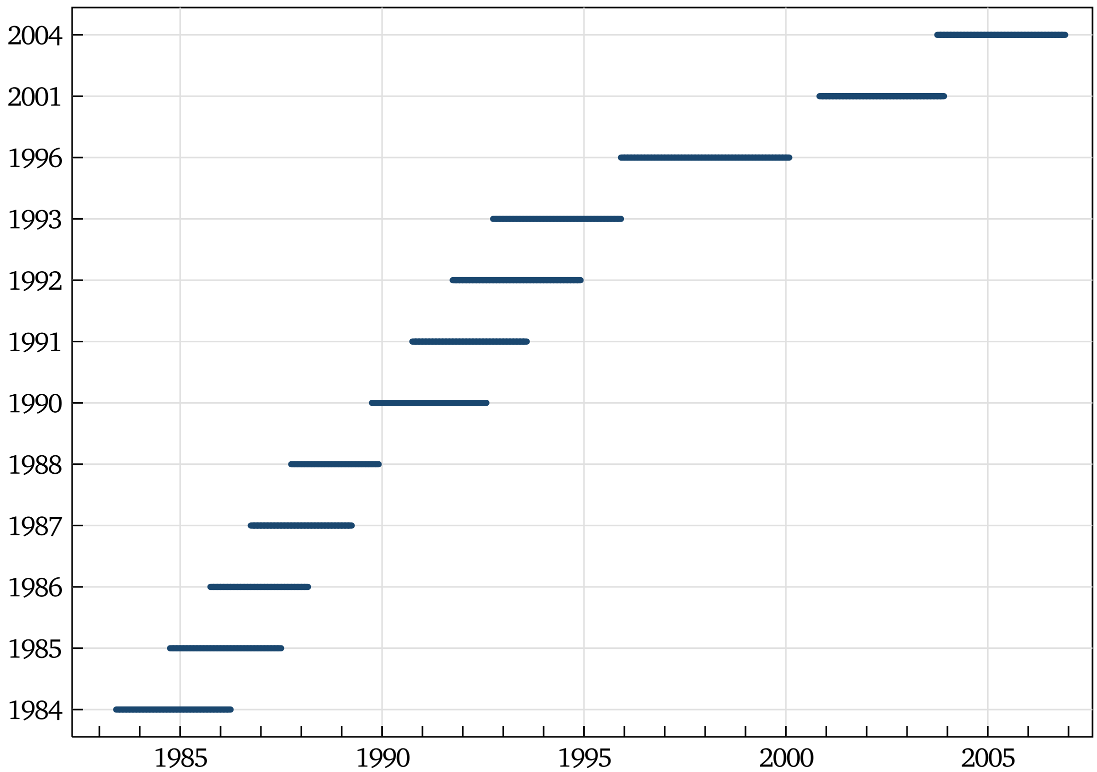

Figure 4.1. SIPP Panel Coverage, 1983–2006[a]

Source: Author's tabulations using SIPP microdata for 1983:6–2006:12.

a. Vertical axis indexes SIPP panels; horizontal lines indicate time coverage of panel.

Percent of population

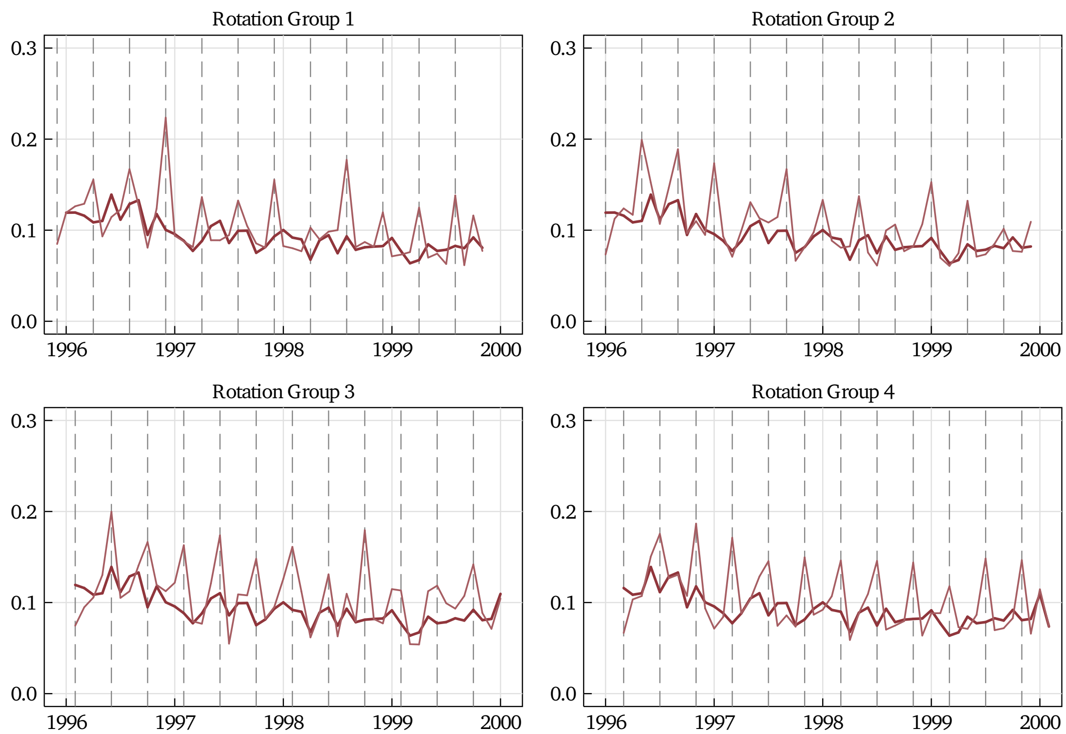

Figure 4.2. The Seam Effect in Separations to Unemployment, 1996 Panel[a]

Source: Author's calculations using SIPP microdata from the 1996 panel.

a. Depicts the flow of persons as a share of the population calculated for each rotation group (thin line) and the seam-corrected average (thick line). Vertical dashed lines indicate wave boundaries.

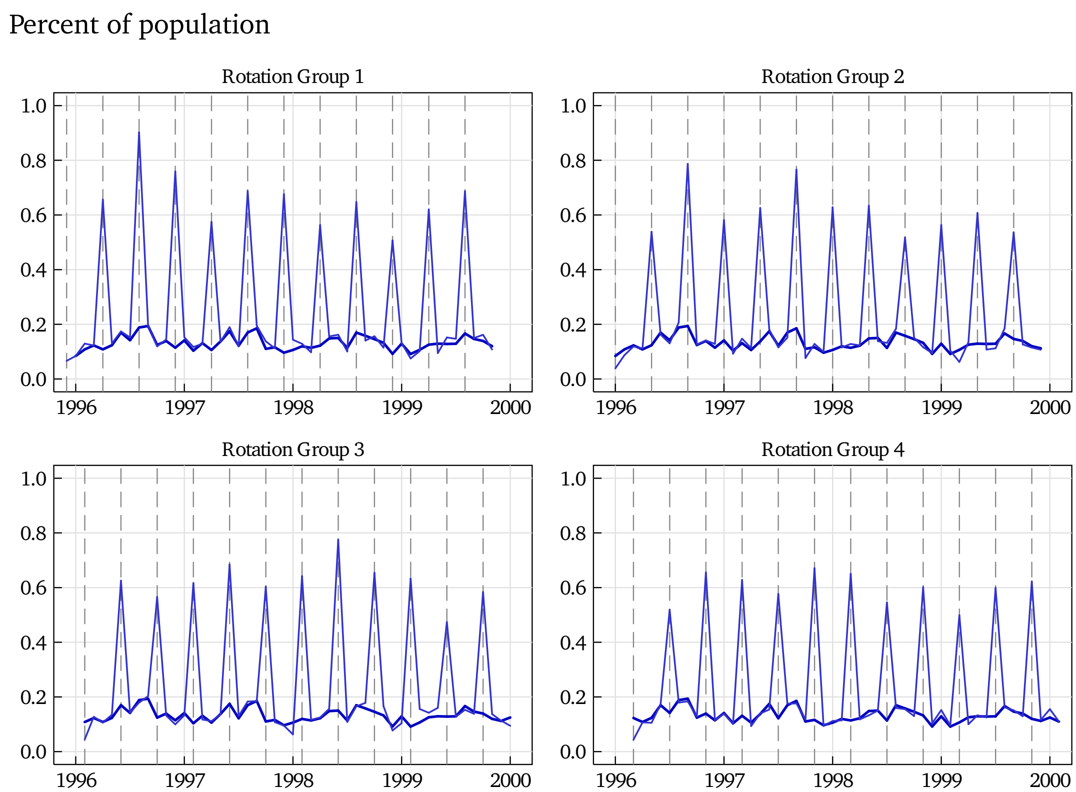

Figure 4.3. The Seam Effect in Job to Job Transitions, 1996 Panel[a]

Source: Author's calculations using SIPP microdata from the 1996 panel.

a. Depicts the flow of persons as a share of the population calculated for each rotation group (thin line) and the seam-corrected average (thick line). Vertical dashed lines indicate wave boundaries.

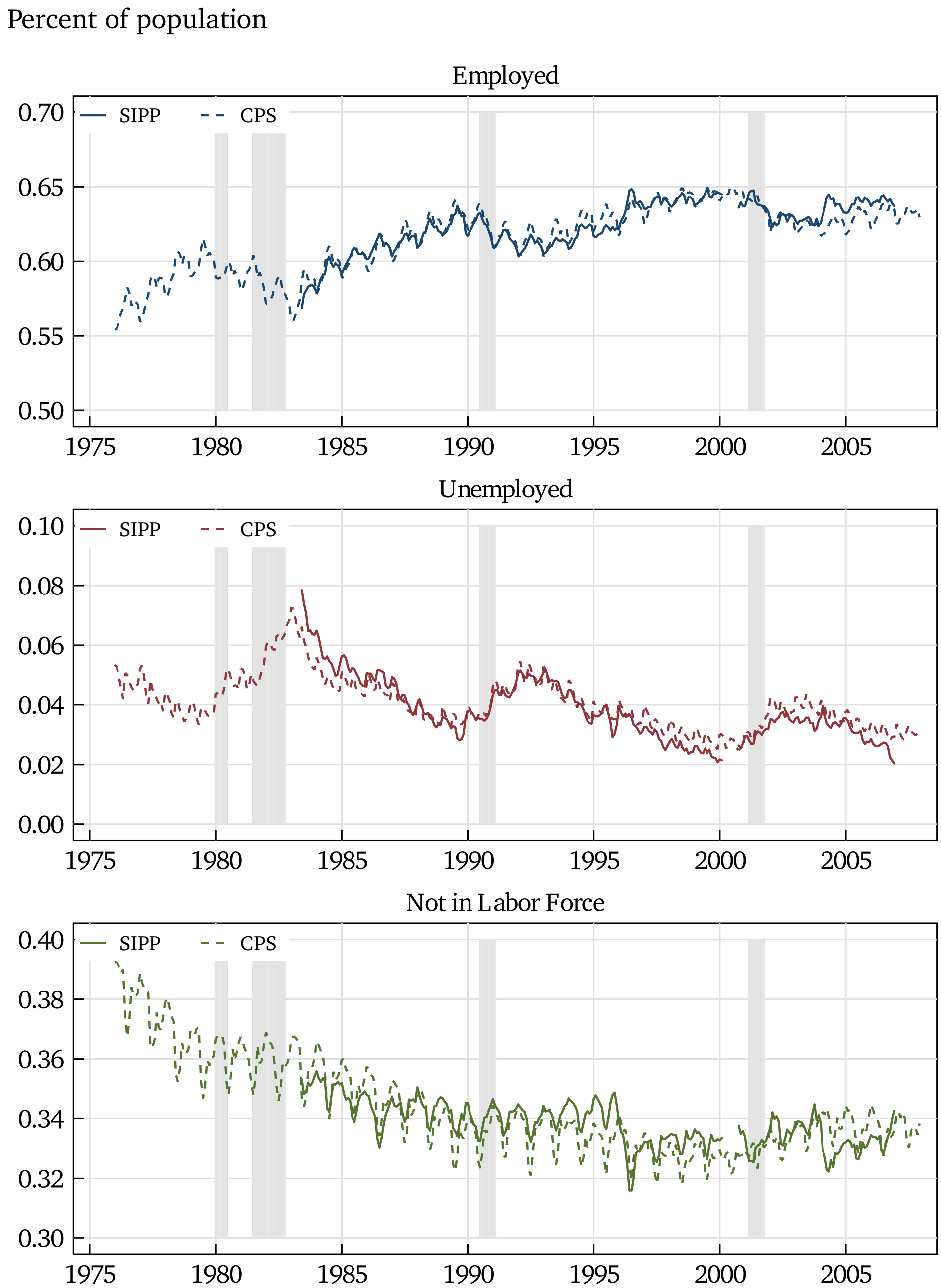

Figure 4.4. Comparison of SIPP and CPS Labor Force Stocks, 1976–2007[a]

Source: Author's calculations using SIPP microdata for 1983:6–2006:12 and CPS data from chapter 3.

a. Data are not seasonally adjusted.

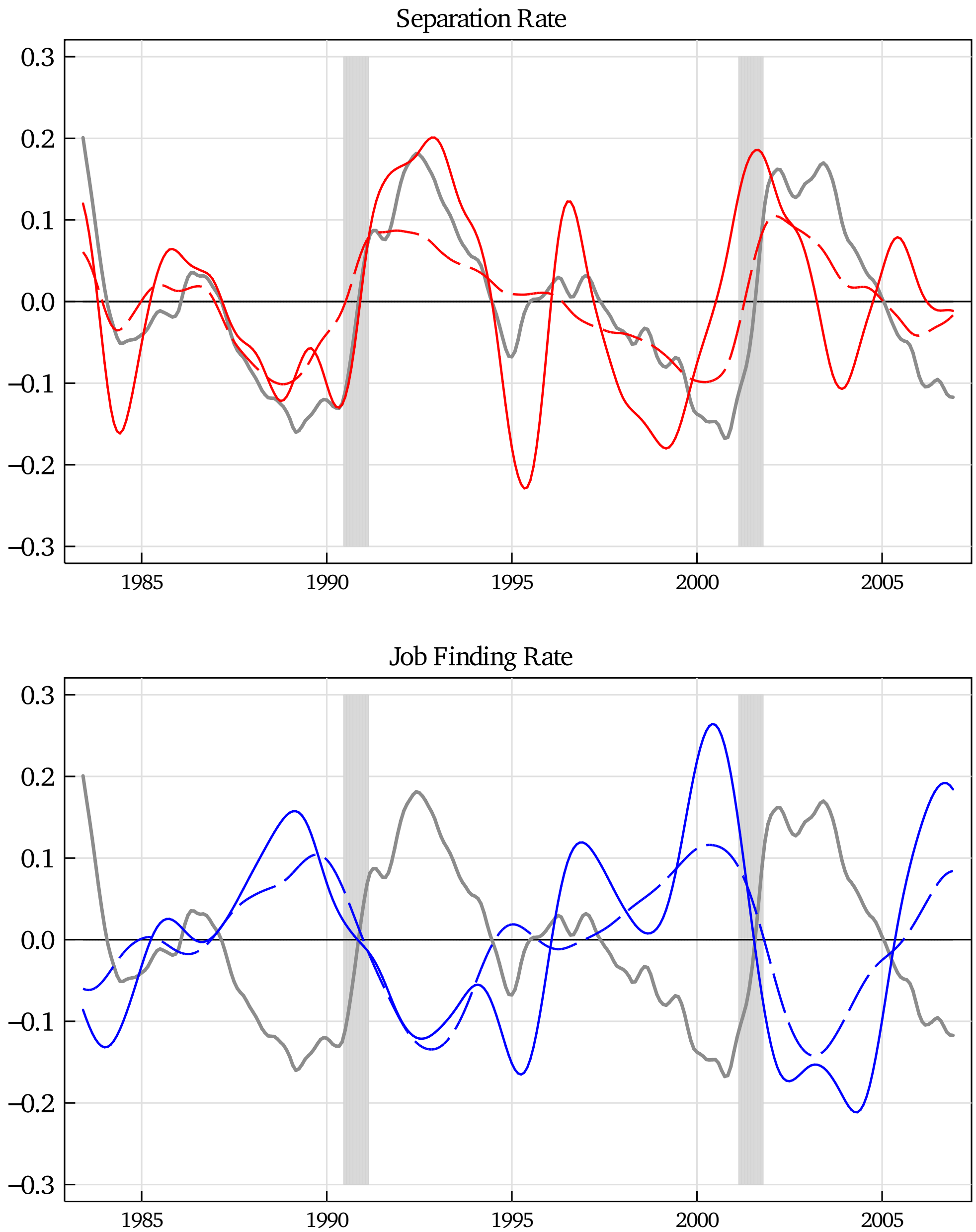

Figure 4.5. Cyclical Component of SIPP and CPS Hazard Rates, 1983–2006[a]

Source: Author's calculations using SIPP microdata for 1983:6–2006:12 and CPS data from chapter 3.

a. Cyclical component estimated using equation 4.8. Gray line is cyclical component of unemployment rate. Thin solid line is SIPP data; dashed line is CPS data. Shaded regions indicate recessions as dated by the National Bureau of Economic Research (NBER).

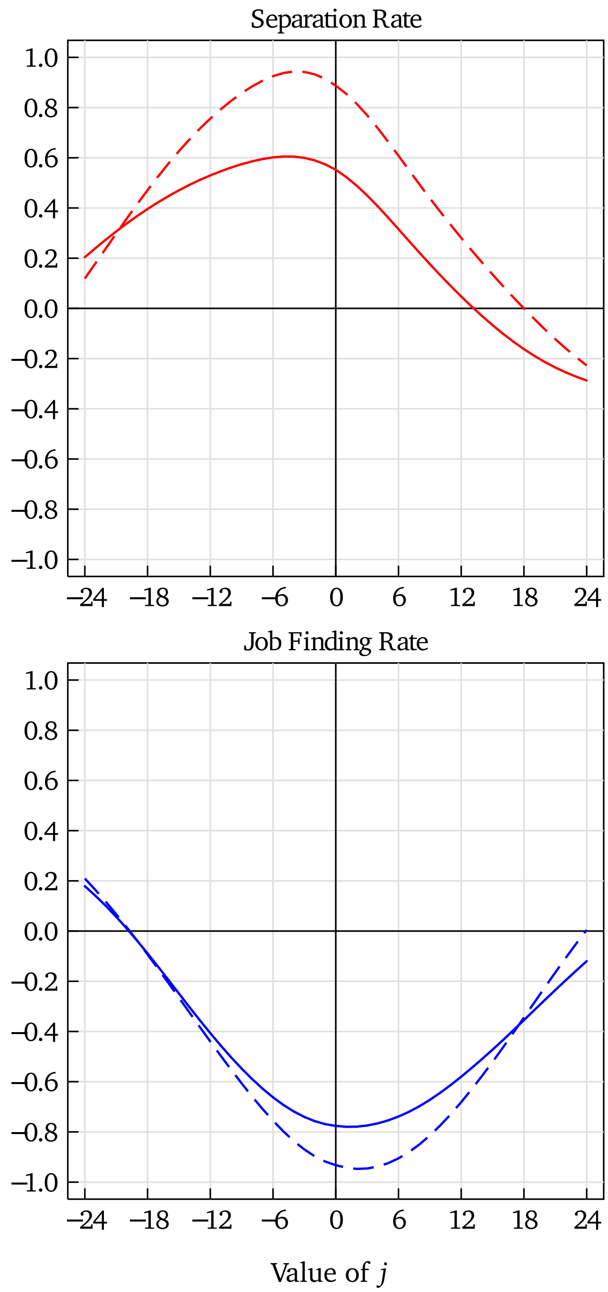

Figure 4.6. Cross-Correlation of Cyclical Components of SIPP and CPS Hazard Rates with Unemployment, 1983–2006[a]

Source: Author's calculations using SIPP and CPS microdata for 1983:6–2006:12 and CPS data from chapter 3.

a. Correlation of $\widehat{\psi}_t^{UR}$ with $\widehat{\psi}_{t+j}^{hr}$, where hr is the appropriate hazard rate. Cyclical component estimated using equation 4.8. Solid line is SIPP data; dashed line is CPS data.

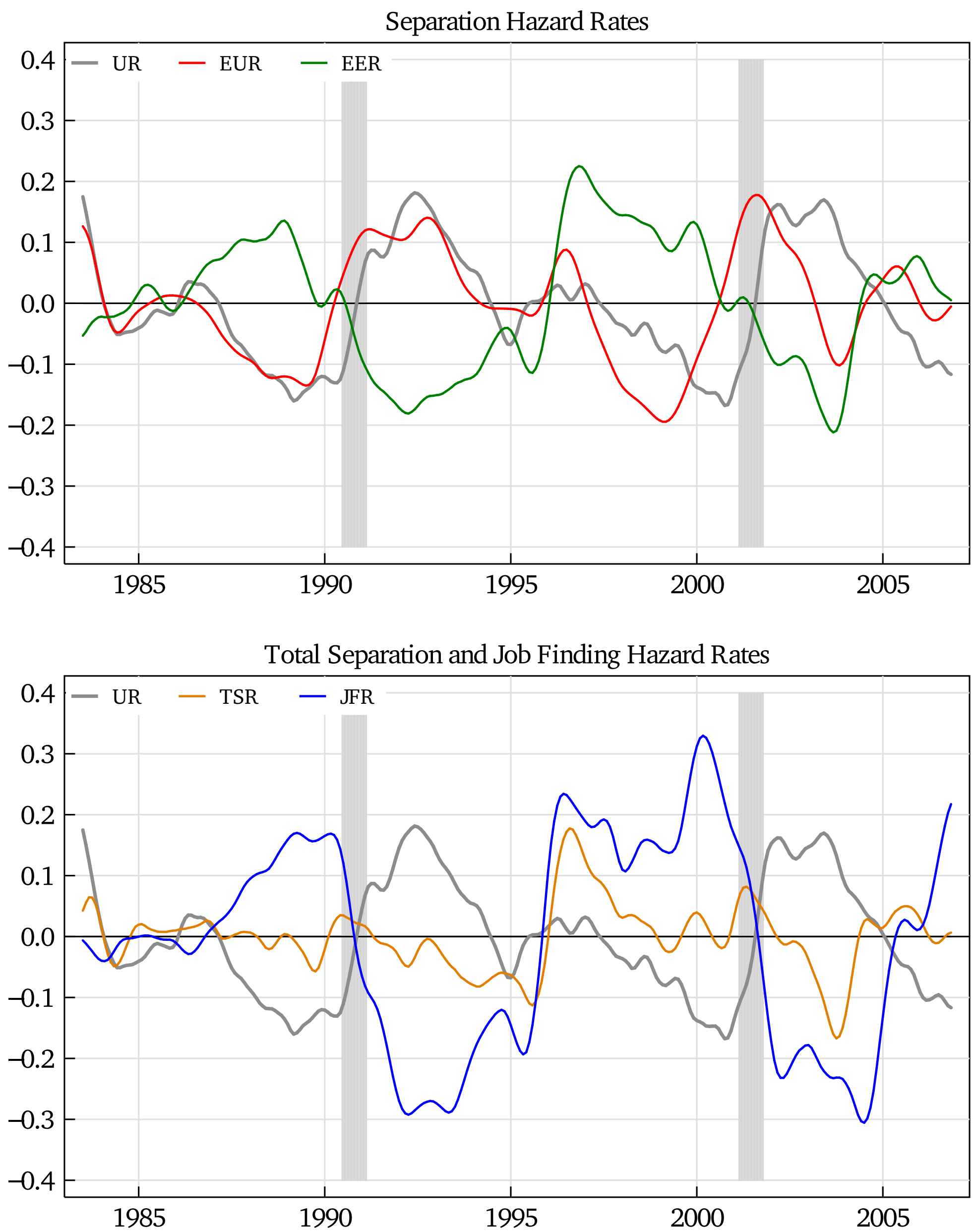

Figure 4.7. Cyclical Component of Weekly Hazard Rates, 1983–2006[a]

Source: Author's calculations using BLS data and SIPP microdata for 1983:7–2006:11.

a. Monthly average of weekly hazard rates. Cyclical component estimated using equation 4.8. Shaded areas indicate recessions as dated by the NBER.

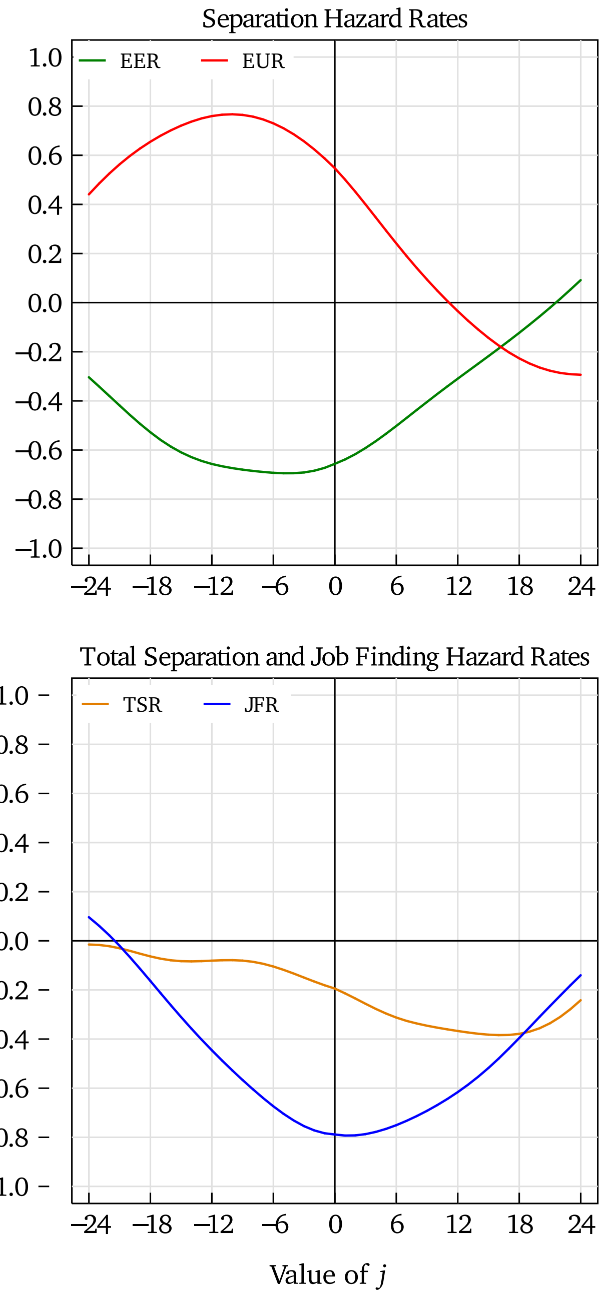

Figure 4.8. Cross-Correlations of Weekly Hazard Rates with Unemployment, CPS, 1983–2006[a]

Source: Author's calculations using BLS data and SIPP microdata for 1983:7–2006:11.

a. Monthly average of weekly hazard rates. Correlation of $\widehat{\psi}_t^{UR}$ with $\widehat{\psi}_{t+j}^{hr}$, where hr is the appropriate hazard rate. Cyclical component estimated using equation 4.8.

Appendix A

Variance Estimation

The SIPP is a multistage stratified survey and, accordingly, estimating the variance requires special consideration. Ignoring the survey design and assuming that observations are selected under simple random sampling understates the true variance.

The SIPP microdata include variables that identify the stratum and primary sampling unit (PSU) from which a person was selected.[1] Because assignment of households to rotation groups is random, the strata from different rotation groups can be thought of as separate strata.

As in the text, let $p = 1, 2, \ldots, 12$ index SIPP panels and $r \in \{1, 2, 3, 4\}$ index the rotation group within a SIPP panel. An individual rotation group is uniquely identified by pr. In month t there are observations from P_t panels, each with R_{pt} rotation groups. For the variance estimation, let $h = 1, 2, \ldots, L_{pr}$ index strata and $i = 1, 2, \ldots, n_{prh}$ index PSUs within rotation group pr. Finally, let $j = 1, 2, \ldots, m_{prhit}$ index persons from rotation group pr within stratum h, PSU i in month t.

The variance estimator for the population ratio (2.2) is

$$\text{(A.1)} \quad \widehat{V}\left(\widehat{T}_t^{\mathtt{IJ}}\right) = \frac{1}{V\left(\widehat{\mathtt{IJ}}_t^{2}\right)} \left\{\widehat{V}\left(\widehat{\mathtt{IJ}}_t^{*}\right) - 2\,\widehat{T}_t^{\mathtt{IJ}} \mathrm{cov}\left(\widehat{\mathtt{IJ}}_t^{*}, \widehat{\mathtt{IJ}}_t\right) + \left(\widehat{T}_t^{\mathtt{IJ}}\right)^2 \widehat{V}\left(\widehat{\mathtt{IJ}}_t\right)\right\}.$$

1. The original PSU and strata codes are not included in the SIPP public use data to maintain confidentiality. Instead, sets of PSUs are combined across strata to produce variance units and variance strata that may be treated as PSUs and strata for variance estimation. See Westat (2001), p. 7–2.

where the survey variance estimator $\widehat{V}\left(\widehat{Y}_t\right)$ is

$$\widehat{V}\left(\widehat{Y}_t\right) = \sum_{p=1}^{P_t}\sum_{r=1}^{R_{pt}}\sum_{h=1}^{L_{pr}} \frac{n_{prh}}{n_{prh}-1} \sum_{i=1}^{n_{prh}} (z_{prhit} - \bar{z}_{prht})^2 , \tag{A.2}$$

and where $z_{prhit} = \sum_{j=1}^{m_{prhit}} w_{prhijt}\, y_{prhijt}$ is the population estimate for stratum h, PSU i in month t and $\bar{z}_{prht} = \frac{1}{n_{prh}} \sum_{i=1}^{m_{prht}} z_{prhit}$ is the mean estimate over stratum h. Finally, the survey covariance estimator is

$$\widehat{\text{cov}}\left(\widehat{Y}_t, \widehat{X}_t\right) = \sum_{p=1}^{P_t}\sum_{r=1}^{R_{pt}}\sum_{h=1}^{L_{pr}} \frac{n_{prh}}{n_{prh}-1} \sum_{i=1}^{n_{prh}} (z^x_{prhit} - \bar{z}^x_{prht})(z^y_{prhit} - \bar{z}^y_{prht}) . \tag{A.3}$$

The 1984–1991 SIPP panels each have 72 variance strata, divided equally among the 4 rotation groups. The 1992 and 1993 panels have 99 strata, the 1996 and 2001 panels have 105, and the 2004 panel has 114. Taken together the pooled SIPP panels contain 1,025 variance strata, each with 2 variance PSUs per stratum.

Appendix B

Structural Time Series Model

The structural time series model for the natural logarithm of each series, denoted y_t, is

$$y_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t, \tag{B.1}$$

where μ_t is the trend, ψ_t the cyclical, γ_t the seasonal, and ε_t the irregular component.

I model the trend component as a smooth first-order local linear trend:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \tag{B.2}$$

$$\Delta\beta_t = \zeta_t, \tag{B.3}$$

where $\Delta = (1 - L)$ and L is the lag operator. The disturbances η_t and ζ_t are independent and identically distributed (i. i. d.) normal random variables with mean zero and variances σ_η^2 and σ_ζ^2.

The cyclical component is modeled as a second-order stochastic cycle with frequency λ, where[1]

$$\begin{bmatrix} \psi_t^{(j)} \\ \psi_t^{*(j)} \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda & \sin\lambda \\ -\sin\lambda & \cos\lambda \end{bmatrix} \begin{bmatrix} \psi_{t-1}^{(j)} \\ \psi_{t-1}^{*(j)} \end{bmatrix} + \begin{bmatrix} \psi_t^{(j-1)} \\ \psi_t^{*(j-1)} \end{bmatrix} \tag{B.4}$$

for $j = 1, 2$ and $\psi_t^{(0)} = \kappa_t$ and $\psi_t^{*(0)} = \kappa_t^*$. The disturbances κ_t and κ_t^* are i. i. d. normal each with mean zero and variance σ_κ^2. Note that for $j = 1$ and $\rho = 1$

1. Harvey and Trimbur (2003) find that, in practice, a second-order cycle provides a good approximation of the gain function of the Baxter-King (BK) bandpass filter.

equation B.4 reduces to a deterministic cycle

$$\psi_t = \psi_0 \cos \lambda t + \psi_0^* \sin \lambda t,$$

where ψ_0 and ψ_0^* are i. i. d. zero-mean random variables with variance σ_ψ^2.

The stochastic seasonal component is constructed so that the s seasonal effects sum to zero in expectation. This is modeled as

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t, \tag{B.5}$$

where $\omega_t \sim N(0, \sigma_\omega^2)$. Finally, the irregular component ε_t is i. i. d. normal with zero mean and variance σ_ε^2. All disturbances are mutually uncorrelated.

The model given by equations B.1–B.5 is represented by the state space system relating observed data y_t to the unobserved state vector $\boldsymbol{\alpha}_t$ through a measurement vector $\mathbf{z}$:

$$y_t = \mathbf{z}'\boldsymbol{\alpha}_t + \varepsilon_t \tag{B.6}$$

$$\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t. \tag{B.7}$$

The unobserved state evolves according to a first-order Markov process with transition matrix $\mathbf{T}$. The state equation (B.7) is

$$\begin{bmatrix} \mu_t \\ \beta_t \\ \psi_t \\ \psi_t^* \\ \gamma_{t-1} \\ \gamma_{t-2} \\ \vdots \\ \gamma_{t-s+2} \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{\text{trend}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_{\text{cycle}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{T}_{\text{seasonal}} \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \psi_{t-1} \\ \psi_{t-1}^* \\ \gamma_{t-2} \\ \gamma_{t-3} \\ \vdots \\ \gamma_{t-s+1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \\ \kappa_t \\ \kappa_t^* \\ \omega_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{B.8}$$

where

$$\mathbf{T}_{\text{trend}} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{T}_{\text{cycle}} = \begin{bmatrix} \rho\cos\lambda & \rho\sin\lambda \\ -\rho\sin\lambda & \rho\cos\lambda \end{bmatrix}$$

$$\underset{(s-1 \times s-1)}{\mathbf{T}_{\text{seasonal}}} = \begin{bmatrix} -1 & -1 & \dots & -1 & -1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ & & \vdots & & \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

This system represents a system with a first-order cycle. The extension to second-order cycles is straightforward.

The state vector enters the measurement equation by the $(4{+}s{-}1 \times 1)$ vector

$$z = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & \dots & 0 \end{bmatrix}'. \tag{B.9}$$

The unknown parameters σ^2_ε, σ^2_η, ρ, λ, σ^2_κ, and σ^2_ω are estimated by maximum likelihood using the Kalman filter. For consistency across all series, I fix the variance of the trend so as to reproduce the Hodrick-Prescott (HP) trend.[2] This variance is $\sigma^2_\zeta = \sigma^2_\varepsilon/129{,}600$.[3] The cycle frequency λ is fixed at sixty months; this corresponds roughly with the center of Burns and Mitchell (1946)'s period of business cycle frequencies. With these restrictions, the estimated trend and cyclical components correspond to a HP lowpass filtered trend and a BK bandpass filtered cyclical component.

2. Harvey and Jaeger (1993) show that the HP trend can be replicated in a structural time series model by a smooth local linear trend with signal-to-noise ratio equal to the inverse of the HP smoothing parameter.

3. Ravn and Uhlig (2002) find the optimal HP smoothing parameter for monthly data is $129{,}600$.

Appendix C

Correcting for Margin Error

Because of the CPS's rotating sample design, at most 75 percent of observations can be matched across succeeding months. The simplest approach is to assume that the unmatched observations are simply missing at random; calculations are performed on the population of matched observations. This assumption has been shown to be a poor one.[1] In particular, the MAR correction significantly undercounts the unemployed.

The conditional MAR model is a simple but powerful extension of the MAR model. Given the timing convention for flows, a person's month t labor force status is always observed, even if the previous month's status is unknown. The MAR model throws this information away. Similar to the corrections of Abowd and Zellner (1985) and Fujita and Ramey (2006), the conditional MAR correction makes use of partially-classified observations. In particular, it assumes that a person missing in month $t-1$ with status $\mathtt{j}$ in month t is drawn randomly from the population of persons with status $\mathtt{j}$ in month t.[2] That is, a person is missing at random conditional on having status $\mathtt{j}$ in month t.

1. See Abowd and Zellner (1985).
2. Lowercase monospaced letters indicate an individual while uppercase monospaced letters indicate a population aggregate.

C.1 Conditional Missing-at-Random Model

The BLS performs much of its second-stage analysis separately by demographic group.[3] In particular, the distinction between male and female and between white and nonwhite are most important. In chapter 2 I adjust for margin error separately by these 4 sex-race groups.

Let $\mathtt{IJ}_{srt}$ be the total number of persons with sex $s \in \{M, F\}$ and race $r \in \{W, NW\}$ who had labor force status $\mathtt{I}$ in month $t-1$ and status $\mathtt{J}$ in month t. Let $\mathtt{MJ}_{srt}$ be the number of persons with missing labor force status in month $t-1$ and status $\mathtt{J}$ in month t. The ratio

$$R^{\mathtt{J}}_{srt} = \frac{\mathtt{EJ}_{srt} + \mathtt{NJ}_{srt} + \mathtt{UJ}_{srt}}{\mathtt{EJ}_{srt} + \mathtt{NJ}_{srt} + \mathtt{UJ}_{srt} + \mathtt{MJ}_{srt}} \tag{C.1}$$

is the number of observed transitions into labor force status $\mathtt{J}$ (flows into $\mathtt{J}$) relative to the total number persons who had labor force status $\mathtt{J}$ in month t (stock of $\mathtt{J}$). The MAR correction normalizes the entire population to the sum of all observed transitions in each labor force status:

$$R^{MAR}_{srt} = \frac{\mathtt{IJ}_{srt}}{\sum_{\mathtt{i}\in\{\mathtt{E,N,U}\}} \sum_{\mathtt{j}\in\{\mathtt{E,N,U}\}} \mathtt{ij}_{srt}}. \tag{C.2}$$

Define the margin error–adjusted $\mathtt{IJ}$ flow, denoted with a tilde, for sex s and race r in month t as

$$\widetilde{\mathtt{IJ}}_{srt} = \frac{\mathtt{IJ}_{srt}}{R^{\mathtt{J}}_{srt}}. \tag{C.3}$$

The table below, reporting the average of monthly adjustment factors over 1976–2007, shows that conditioning on month t labor force status makes a large difference. Averaging across demographic groups, the MAR model (equation C.2) would inflate each measured flow by a factor of 1.43. The conditional MAR model, in contrast, inflates flows ending in employment by 1.43 but flows ending in unemployment by 1.51. The conditional MAR model identifies about 6 percent more flows into unemployment than the unconditional model.

3. Bureau of Labor Statistics (2002).

Demographic group	MAR	Status in month t: E	U	N
Male, white	1.4235	1.4245	1.4995	1.4081
Male, nonwhite	1.4569	1.4583	1.5303	1.4371
Female, white	1.4178	1.4215	1.5161	1.4075
Female, nonwhite	1.4463	1.4485	1.5393	1.4314

The population margin error–adjusted IJ flow in month t is the sum over sex and race categories:

$$\widetilde{\mathtt{IJ}}_t = \sum_s \sum_r \widetilde{\mathtt{IJ}}_{srt}. \tag{C.4}$$

C.2 Conditioning on Geographic Mobility

In chapter 3 I condition on geographic mobility, in addition to conditioning on sex and race as in the previous section.

Let $m = 0$ for persons that do not move and $m = 1$ for movers. Let $\mathtt{IJ}_{srmt}$ be the number of persons with sex $s \in \{M, F\}$ and race $r \in \{W, NW\}$ who had labor force status i in month $t-1$ and status j in month t. Let $\mathtt{MJ}_{srmt}$ be the number of persons with missing labor force status in month $t-1$ and status j in month t. The ratio

$$R^{\mathtt{J}}_{srmt} = \frac{\mathtt{EJ}_{srmt} + \mathtt{NJ}_{srmt} + \mathtt{UJ}_{srmt}}{\mathtt{EJ}_{srmt} + \mathtt{NJ}_{srmt} + \mathtt{UJ}_{srmt} + \mathtt{MJ}_{srmt}} \tag{C.5}$$

is the number of observed transitions into labor force status J relative to the total number persons who had labor force status J in month t.

Define the margin error–adjusted IJ flow, denoted with a tilde, for mobility m, sex s, and race r in month t as

$$\widetilde{\mathtt{IJ}}_{srmt} = \frac{\mathtt{IJ}_{srmt}}{R^{\mathtt{J}}_{srmt}}. \tag{C.6}$$

The aggregate margin error–adjusted IJ flow in month t is the sum over the 6 mobility-sex-race cells:

$$\widetilde{\mathtt{IJ}}_t = \sum_m \sum_s \sum_r \widetilde{\mathtt{IJ}}_{srmt}. \tag{C.7}$$

Appendix D

LPD Construction

This appendix provides details about the construction of the Longitudinal Population Database (LPD). The database is compiled in two stages. In the first stage the raw data for each month are imported into a statistical program and processed to ensure that all variables are longitudinally consistent across all marks. In the second stage the processed monthly files are appended together to create a longitudinal data set. The entire data set is then processed to properly identify addresses, households and household changes, and individuals using all longitudinal information.

D.1 Stage I: Raw Data

In stage I, the monthly data files are processed individually. Each is imported into a statistical program and then processed to create longitudinally-consistent variables.

The monthly public-use CPS microdata flat files are downloaded from the Census Bureau and the CPS data repository at the NBER. The Census Bureau web site hosts the microdata files for 1992 to the present. Data before 1992 come from the NBER, which maintains copies of the data files for 1976 to the present.

The variable layout and definitions in the microdata files change 17 times over 1976–2006. Each different version of the layout and definition is called a "mark." Many of the variable locations and definitions remain the same across marks, however a change in any one variable constitutes a new mark. The table below lists the 18 marks and the months they span.

Mark number	*Start date*	*End date*	*No. of months*
0	Jan 1976	Dec 1977	24
1	Jan 1978	Dec 1981	48
2	Jan 1982	Dec 1982	12
3	Jan 1983	Dec 1983	12
4	Jan 1984	Jun 1985	18
5	Jul 1985	Dec 1985	6
6	Jan 1986	Dec 1988	36
7	Jan 1989	Dec 1991	36
8	Jan 1992	Dec 1993	24
9	Jan 1994	Mar 1994	3
10	Apr 1994	May 1995	14
11	Jun 1995	Aug 1995	3
12	Sep 1995	Dec 1997	28
13	Jan 1998	Dec 2002	60
14	Jan 2003	Apr 2004	16
15	May 2004	Jul 2005	15
16	Aug 2005	Dec 2006	17
17	Jan 2007	Dec 2007	12

D.2 Stage I: Data Dictionaries

To construct a longitudinal database, all variable names and definitions must be the same across all marks. Because many change from mark to mark, I compare the 18 data definition files and create a set of universal variable names and definitions that are consistent across all marks. Table D.1 reports the universal variable names and definitions.

I then create dictionaries for each mark that correspond to the universal definitions. Jean Roth at the NBER provides data dictionaries for marks 7–16, however they do not conform to the universal definitions.[1] I modify these dictionaries to maintain longitudinal consistency.

1. `http://www.nber.org/cps-basic/`.

D.3 Stage I: Longitudinal Consistency

This subsection describes how the LPD's variables are created from CPS variables to ensure longitudinal consistency across all marks. Variable names are set in monospaced type; those in uppercase identify variables from the CPS while those in lowercase are LPD variables.

D.3.1 Survey Date

All observations in the CPS contain the 2-digit month of the survey (`HRMONTH`) and some measure of the year (`HRYEAR`). For marks 1–8 the CPS reports only the last digit of the survey year, while marks 9–12 report the last 2 digits. All other marks include the 4-digit year. The LPD variable year is constructed from `HRYEAR` to report the full 4-digit year of survey. The information on month is unaltered.

D.3.2 Interview Status

The CPS reports the status of each interview in the variable `HRINTSTA`. The interview status can take on 4 values: completed interview, type A noninterview, type B noninterview, and type C noninterview. The LPD variable `INTSTAT` reports this code for marks 9–16. Prior to mark 9 the CPS classifies interview status into only 3 categories, combining type B and type C noninterviews into one category. For these marks the type B and type C noninterviews are separated using supplementary information.

D.3.3 State

The CPS records the U.S. state of the address using two different code systems. Marks 7–16 report the state using both the Federal Information Processing System (FIPS) code (`GESTFIPS`) and the Census Bureau state code (`GESTCEN`). Prior to mark 6, the CPS reports only `GESTCEN`. The LPD variable `STATE` contains the FIPS state code for the address. The concordance between Census state codes and FIPS state codes is below.

State	FIPS code	Census code	State	FIPS code	Census code	State	FIPS code	Census code
AK	02	94	KY	21	61	NY	36	21
AL	01	63	LA	22	72	OH	39	31
AR	05	71	MA	25	14	OK	40	73
AZ	04	86	MD	24	52	OR	41	92
CA	06	93	ME	23	11	PA	42	23
CO	08	84	MI	26	34	RI	44	15
CT	09	16	MN	27	41	SC	45	57
DC	11	53	MO	29	43	SD	46	45
DE	10	51	MS	28	64	TN	47	62
FL	12	59	MT	30	81	TX	48	74
GA	13	58	NC	37	56	UT	49	87
HI	15	95	ND	38	44	VA	51	54
IA	19	42	NE	31	46	VT	50	13
ID	16	82	NH	33	12	WA	53	91
IL	17	33	NJ	34	22	WI	55	35
IN	18	32	NM	35	85	WV	54	55
KS	20	47	NV	32	88	WY	56	83

D.3.4 Sex

No changes to the coding are required.

D.3.5 Race

The level of detail for racial classification varies widely across the marks. There are 3 major classification schemes. The most recent marks (14–16) classify race into 21 separate categories (`PRDTRACE`). Marks 7–13 have 5 distinct categories (`PERACE`) and marks 1–6 report only 3: white, black, and other. Thus, to maintain longitudinal consistency, race is recoded into the 3 categories from marks 1–6. Below is a concordance for the two other schemes.

Mark 14–16		*Mark 7–13*	
`PRDTRACE`	`RACE`	`PERACE`	`RACE`
WHITE	WHITE	WHITE	WHITE
BLACK	BLACK	BLACK	BLACK
AMERICAN INDIAN (AI)	OTHER	AMERICAN INDIAN	OTHER
ASIAN	OTHER	ASIAN-PACIFIC ISLANDER	OTHER
HAWAIIAN (HP)	OTHER	OTHER	OTHER
WHITE-BLACK	WHITE		
WHITE-AI	WHITE		
WHITE-ASIAN	WHITE		
WHITE-HP	WHITE		
BLACK-AI	BLACK		
BLACK-ASIAN	BLACK		
BLACK-HP	BLACK		
AI-ASIAN	OTHER		
ASIAN-HP	OTHER		
WHITE-BLACK-AI	WHITE		
WHITE-BLACK-ASIAN	WHITE		
WHITE-AI-ASIAN	WHITE		
WHITE-ASIAN-HP	WHITE		
WHITE-BLACK-AI-ASIAN	WHITE		
2 OR 3 RACES	OTHER		
4 OR 5 RACES	OTHER		

D.3.6 Age

The CPS reports each individual's age as of the end of the reference week (`PEAGE`), topcoded at different years depending on the mark. For most variables the CPS reports information with greater detail as the survey ages, but this is not the case with age: marks 1–4 topcode ages above 99 years old, marks 5–14 topcode ages above 90, and marks 15–16 topcode ages above 80. The LPD variable `AGE` is re-topcoded as 80 for ages 80–84 and as 85 for ages 85 and older.

D.3.7 Marital Status

The CPS classifies marital status (`PEMARITL`) using 3 different schemes. The LPD classifies marital status (`MS`) as either married, widowed/divorced, or never married. The concordance with the CPS data is below.

PEMARITL	MS
Mark 9–16	
MARRIED-SPOUSE PRESENT	MARRIED
MARRIED-SPOUSE ABSENT	MARRIED
WIDOWED	WIDOWED/DIVORCED
DIVORCED	WIDOWED/DIVORCED
SEPARATED	MARRIED
NEVER MARRIED	NEVER MARRIED
Mark 6–8	
MARRIED-CIVILIAN SPOUSE PRESENT	MARRIED
MARRIED-AF SPOUSE PRESENT	MARRIED
MARRIED-SPOUSE ABSENT	MARRIED
WIDOWED	WIDOWED/DIVORCED
DIVORCED	WIDOWED/DIVORCED
SEPARATED	MARRIED
NEVER MARRIED	NEVER MARRIED
Mark 1–5	
MARRIED-CIVILIAN SPOUSE PRESENT	MARRIED
MARRIED-AF SPOUSE PRESENT	MARRIED
MARRIED-SPOUSE ABSENT	MARRIED
WIDOWED OR DIVORCED	WIDOWED/DIVORCED
NEVER MARRIED	NEVER MARRIED

D.3.8 Educational Attainment

As part of the 1994 survey redesign, the CPS changed the education question from a quantitative question about the years of schooling attended to a qualitative question about level of education attained. Jaeger (1997) studied the relationship between the two questions by comparing responses from individuals who answered both versions of the question. The LPD education variable, `EDUC`, is coded using Jaeger's correspondence, reported below.

	Highest grade attended		*Educational*
Category	*Not completed*	*Completed*	*attainment*
High school dropout	0–12	1–11	31–37
High school graduate	n.a.	12	38, 39
Some college	13–16	13–15	40–42
College graduate	17, 18	16–18	43–46

D.3.9 Labor Force Status

The labor force status ultimately reported in the CPS (PEMLR) is a recode based on answers to survey questions. Although the broad classification of labor force status—employed, unemployed, and NILF—is unchanged throughout the history of the CPS, the labor force subclassifications do change. The LPD classifies labor force status into those 3 broad categories as follows:

PEMLR	LFS
Mark 12–16	
EMPLOYED-AT WORK	EMPLOYED
EMPLOYED-ABSENT	EMPLOYED
UNEMPLOYED-ON LAYOFF	UNEMPLOYED
UNEMPLOYED-LOOKING	UNEMPLOYED
NILF-RETIRED	NILF
NILF-DISABLED	NILF
NILF-OTHER	NILF
Mark 6–11	
EMPLOYED-AT WORK	EMPLOYED
EMPLOYED-ABSENT	EMPLOYED
UNEMPLOYED-ON LAYOFF	UNEMPLOYED
UNEMPLOYED-LOOKING	UNEMPLOYED
NILF-WORK W/O PAY	NILF
NILF-UNAVAILABLE	NILF
NILF-OTHER	NILF
Mark 1–5	
EMPLOYED-AT WORK	EMPLOYED
EMPLOYED-ABSENT	EMPLOYED
UNEMPLOYED-LOOKING	UNEMPLOYED
NILF-HOUSE	NILF
NILF-SCHOOL	NILF
NILF-UNABLE	NILF
NILF-OTHER (INC. RETIRED)	NILF

D.3.10 Industry and Occupation

No changes to the coding are required.

D.4 Stage II: Observation Identifiers

This section describes how the longitudinal units of the LPD are defined and constructed; see chapter 3.

The first step of stage II is to append the processed monthly data files together into a single longitudinal data set. The observations are then sorted chronologically by a unique address identifier to organize them into a time series for each address. The addresses are then processed to identify households and the households processed to identify individuals. The following sections describe how these identifiers are constructed.

D.4.1 Address Identifier

The primary identifier in the CPS is a "unique household identifier" (`HRHHID`). All observations in the CPS have a `HRHHID`. This variable does not, however, identify households nor is it unique, either locally (within a single month) or globally (both within and across months). More precisely, it is a partial *address* identifier that, together with other variables, uniquely identifies an address. For marks 1–11 `HRHHID` is a 12-digit number, but for marks 12–16 it increases to 15 digits. All 12-digit `HRHHID`s are padded to 15 digits by adding 3 leading zeros.

For mark 10 and later, an address is uniquely identified by `HRHHID` and 2 other variables: the sample identifier (`HRSAMPLE`) and serial suffix (`HRSERSUF`). Concatenating these three variables creates a 19-digit, globally-unique address identifier, `AID`. An address observation unit (AOU) is defined as all observations with the same `AID`.

Unfortunately, the 2 additional variables needed to create `AID` do not exist in marks 1–10. This problem manifests itself when the data from all months are combined into one longitudinal data set. Because an address is not uniquely identified across months, observations from several different addresses will have the same `HRHHID`. This collection of all observations from a single household identifier is called a `HRHHID` group. The figure below illustrates the problem. The top row displays fictional data for an address uniquely identified by `HRHHID` (and the supplemental variables), such as one from marks 11–16. The cell values are the address's month in sample (MIS). The bottom two rows show fictional data for 2 `HRHHID` groups from marks 1–10. Each `HRHHID` has observations for many more months than is possible under the CPS survey design.

	Month																									
HRHHID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**24**	**26**
…0026						1	2	3	4									5	6	7	8					
…5923	1	2	3	4	1	2	3	4	1	2	3	4	5	6	7	8	5	6	7	8	5	6	7	8	1	2
…8321	7	8	1	2	3	4	1	2	3	4	1	2	3	4	5	6	7	8	5	6	7	8	5	6	7	8

Because each address is surveyed according to a defined rotation pattern, there is a unique relationship between the survey date and an address's MIS within an `HRHHID` group. For example, an address that enters the CPS sample at calendar month 6 can have at most 4 interviews at months 6–9 and 4 more interviews at months 18–21 (for example, the `HRHHID` ending with 0026 in the figure above). If the data from the `HRHHID` group at month 18 does not have MIS = 5, then it must be from a different address. I have written an algorithm that exploits this relationship to uniquely identify individual addresses within a `HRHHID` group.

The figure below illustrates how the observations from the 2 fictional `HRHHID` groups are separated into different addresses under the address algorithm.

	Month																									
HRHHID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**24**	**26**
…5923	1	2	3	4	1	2	3	4	1	2	3	4	5	6	7	8	5	6	7	8	5	6	7	8	1	2
…5923A	1	2	3	4									5	6	7	8										
…5923B					1	2	3	4									5	6	7	8						
…5923C									1	2	3	4									5	6	7	8		
…5923D																									1	2
…8321	7	8	1	2	3	4	1	2	3	4	1	2	3	4	5	6	7	8	5	6	7	8	5	6	7	8
…321A	7	8																								
…8321B			1	2	3	4									5	6	7	8								
…8321C							1	2	3	4									5	6	7	8				
…8321D											1	2	3	4									5	6	7	8

About 17 percent of AOUs have no completed interviews. Of these over 85 percent consist exclusively of type B or type C noninterviews. These AOUs are discarded because they contribute no data, longitudinal or otherwise.[2] This is different from discarding a single interview that has no data in a particular month; these AOUs are never eligible for interview during their entire CPS history. Line 4 in table 3.1 depicts an example of an AOU that would be discarded. The remaining AOUs with no completed interviews consist of all type A noninterviews. These AOUs remain in the sample because these addresses contain households that could have been interviewed.

D.4.2 Household Identifier

After creating a unique address identifier, all addresses are processed to identify unique households. Section 3.2.2 describes the 4 ways a household change can occur within an AOU. I have written an algorithm that identifies these household changes and creates a unique household identifier.

The CPS records the number of households that occupy an address during its 8-interview history. Each time a new household is identified at an address, the household number (`HUHHNUM`) is incremented. There may be up to 8 different households at an address. For addresses without a noninterview, the household number (`HNUM`) is given by `HUHHNUM`. This correctly identifies type H1 household changes (no intervening vacancy).

Addresses with noninterviews require special processing to create the correct household number. The CPS does not change `HUHHNUM` following a type B or type C noninterview.[3] Therefore, all observations after a type B or type C noninterview are assigned to the same household when they must be from a different household. The

2. If addresses are ineligible at random, excluding them does not bias the sample. If, however, a disproportionate number of addresses selected were located, for example, in a poor inner-city and had been condemned, then excluding these AOUs could bias the estimate. Comparing the distribution across states of AOUs with noninterviews against those without noninterviews reveals no substantive differences.

3. Type B and type C noninterviews indicate the address is ineligible for interview that month, implying that the future (previous) occupants are not the same as the previous (future) occupants. Had the same occupants simply been unavailable that month, the interview would have been recorded as a type A noninterview.

household algorithm correctly identifies the remaining 3 types of household change.

For each address, the household algorithm examines all observations with the same household number in chronological order. This is the largest group of observations that could be from the same household. When it encounters a type B or type C noninterview it does the following:

1. if there are no valid (completed interview or type A noninterview) observations in the past, the current observation is dropped;

2. if there are valid observations in the past but none in the future, the current observation is dropped; or

3. if there are valid observations in the past and valid observations in the future, the current observation is dropped and all future observations from this address are assigned the next `HNUM`.

The algorithm continues until all observations from an address with the same household number have been processed. It then repeats for the next address. Appending `HNUM` to `AID` creates a 20-digit, globally-unique household identifier, `HID`. A household observation unit (HOU) is defined as all observations with the same `HID`.

D.4.3 Person Identifier

The CPS identifies individuals within a household by their line number on the survey response sheet (`PULINENO`). An individual retains the same line number for each month in the survey. Appending the 2-digit line number to hid creates a 22-digit, globally-unique person identifier, `PID`. When `PULINENO` is less than 2 digits, a leading zero is added. A person observation unit (POU) is defined as all observations with the same `PID`.

This procedure does not work, however, for persons at an address with a type A noninterview.[4] Because no survey is performed that month, no information on the number of persons at the address is collected. Thus, because no line number exists that month, the longitudinal continuity of all POUs at the address is interrupted.

4. Note that at this point in processing all type B and type C noninterviews have been dropped. The term noninterview thus refers only to type A noninterviews.

The third algorithm I create processes households for noninterviews and generates line numbers for persons living at the address in months with noninterviews. For each household, the person algorithm searches in chronological order for a noninterview. When it finds a noninterview it does the following:

1. if there is a valid observation in the previous month, the current month's observation is duplicated for each person at the address during the previous month;
2. if there are no valid observations in the past but there is a valid observation in the future, the current month's observation is duplicated for each person at the address during the first future month with a valid observation; or
3. if there are no valid observations in the past or in the future, the current month's observation is given a line number of 1.

The newly-created observations for each person contain the same information as the original address-level observation. That is, they have only information on the address and interview status. They contain no demographic or labor force information. The person algorithm does not currently attempt to impute this missing information.

It is impossible to know, a priori, if the persons who occupy the address in a month with a noninterview are the same as those who occupy the address in the previous or subsequent months. The algorithm assigns line numbers based on the last known observation or, when no previous valid observation exists, on the closest future observation.

Table D.1. LPD Variable Definitions

Variable name	*Description*	*Value*
AID	Unique address identifier	19-digit string
HID	Unique household identifier	20-digit string
PID	Unique person identifier	22-digit string
MONTH	Month of interview	2-digit number
YEAR	Year of interview	4-digit number
MIS	Month in sample	1-digit number [1-8]
INTSTAT	Interview status	1 – Interview 2 – Type A noninterview 3 – Type B noninterview 4 – Type C noninterview
STATE	FIPS state code	2-digit number [1-56]
SEX	Sex	1 – Male 2 – Female
RACE	Race	1 – White 2 – Black 3 – Other
AGE	Age at end of reference week	2-digit number [0-85], topcoded as 80 for ages 80–84 topcoded as 85 for ages 85+
MS	Marital status	1 – Married 2 – Widowed/divorced 3 – Never married
EDUC	Educational attainment	1 – Less than high school graduate 2 – High school graduate 3 – Some college 4 – College graduate
LFS	Labor force status	e – Employed u – Unemployed n – Not in the labor force
IND	Major industry recode	2-digit number [1-23], see CPS data definitions
OCC	Major occupation recode	2-digit number [1-15], see CPS data definitions

Bibliography

Abowd, John and Arnold Zellner, 1985. "Estimating Gross Labor Force Flows." *Journal of Business and Economic Statistics* 3(3): 254–83.

Bartel, Ann P., 1979. "The Migration Decision: What Role Does Job Mobility Play?" *The American Economic Review* 69(5): 775–86.

Baxter, Marianne and Robert G. King, 1999. "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series." *The Review of Economics and Statistics* 81(4): 575–93.

Bils, Mark, Yongsung Chang, and Sun-Bin Kim, 2007. "Comparative Advantage in Cyclical Unemployment." Working Paper 13231, National Bureau of Economic Research.

Blanchard, Olivier-Jean and Peter Diamond, 1990. "The Cyclical Behavior of the Gross Flows of U.S. Workers." *Brookings Papers on Economic Activity* (2): 85–143.

Bleakley, Hoyt, Ann Ferris, and Jeffrey Fuhrer, 1999. "New Data on Worker Flows During Business Cycles." *New England Economic Review* (July/August): 49–76.

Bureau of Labor Statistics, 2002. "Current Population Survey: Design and Methodology." Technical Paper 63RV, Washington, DC: U.S. Census Bureau.

Burns, Arthur F. and Wesley C. Mitchell, 1946. *Measuring Business Cycles.* National Bureau of Economic Research.

Chua, Tin Chiu and Wayne A. Fuller, 1987. "A Model for Multinomial Response Error Applied to Labor Flows." *Journal of the American Statistical Association* 82(397): 46–51.

Cleveland, William S., 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74(368): 829–36.

Cogley, Timothy and James M. Nason, 1995. "Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle Research." *Journal of Economic Dynamics and Control* 19(1–2): 253–78.

Dahmann, Donald C., 1986. "Geographical Mobility Research with Panel Data." *Growth and Change* 17(3): 35–48.

Darby, Michael, John Haltiwanger, and Mark Plant, 1986. "The Ins and Outs of Unemployment: The Ins Win." Working Paper 1997, NBER.

Davis, Stephen J. and John Haltiwanger, 1990. *NBER Macroeconomics Annual*, chapter Gross Job Creation and Destruction: Microeconomic Evidence and Macroeconomic Implications, pp. 123–68. MIT Press.

Davis, Steven J., R. Jason Faberman, John C. Haltiwanger, and Ian Rucker, 2008. "Adjusted Estimates of Worker Flows and Job Openings in JOLTS." Working Paper 14137, National Bureau of Economic Research.

Durbin, James and Siem Jan Koopman, 2001. *Time Series Analysis by State Space Models*. Oxford University Press.

Elsby, Michael W., Ryan Michaels, and Gary Solon, forthcoming. "The Ins and Outs of Cyclical Unemployment." *American Economic Journal: Macroeconomics*.

Faberman, R. Jason, 2005. "Analyzing the JOLTS Hires and Separations Data." In "Proceedings of the Annual Statistical Association," Alexandria, VA: American Statistical Association.

Fallick, Bruce and Charles A. Fleischman, 2004. "Employer-to-Employer Flows in the U.S. Labor Market: The Complete Picture of Gross Worker Flows." Working Paper 2004-34, Federal Reserve Board of Governors.

Farber, Henry S., 1999. *Mobility and Stability: The Dynamics of Job Change in Labor Markets*, volume 3B of *Handbook of Labor Economics*, pp. 2439–83. Amsterdam: Elsevier Science, North-Holland.

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

Fisher, Ronald A., 1915. "Frequency Distribution of the Values of the Correlation Coefficient in Samples of an Indefinitely Large Population." *Biometrika* 10(4): 507–21.

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt, 1998. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." *The Journal of Human Resources* 33(2): 251–99.

Fujita, Shegiru, Christopher J. Nekarda, and Garey Ramey, 2007. "The Cycality of Worker Flows: New Evidence from the SIPP." Working Paper 07-05, Federal Reserve Bank of Philadelphia.

Fujita, Shegiru and Garey Ramey, forthcoming. "The Cyclicality of Separation and Job Finding Rates." *International Economic Review* .

Fujita, Shigeru and Garey Ramey, 2006. "The Cyclicality of Job Loss and Hiring." Working Paper 06-17, Federal Reserve Bank of Philadelphia.

Fujita, Shigeru and Garey Ramey, 2007. "Reassessing the Shimer Facts." Working Paper 07-02, Federal Reserve Bank of Philadelphia.

Gallaway, Lowell E., 1969. "Age and Labor Mobility Patterns." *Southern Economic Journal* 36(2): 171–80.

Gottschalk, Peter and Robert Moffitt, 1999. "Changes in Job Instability and Insecurity Using Monthly Survey Data." *Journal of Labor Economics* 17(4): S91–S126.

Greenwood, Michael J., 1975. "Research on Internal Migration in the United States: A Survey." *Journal of Economic Literature* 13(2): 397–433.

Hagedorn, Marcus and Iourii Manovskii, 2008. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited." *American Economic Review* forthcoming.

Hall, Robert E., 2005. "Employment Efficiency and Sticky Wages: Evidence from Flows in the Labor Market." *Review of Economics and Statistics* 87(3): 397–407.

Hall, Robert E., 2006. *NBER Macroeconomics Annual 2005*, chapter Job Loss, Job Finding, and Unemployment in the U.S. Economy over the Past Fifty Years, pp. 275–310. Cambridge, Mass.: National Bureau of Economic Research.

Harvey, Andrew C., 1989. *Forecasting, Structual Time Series Models, and the Kalman Filter.* Cambridge University Press.

Harvey, Andrew C. and Albert Jaeger, 1993. "Detrending, Stylised Facts and the Business Cycle." *Journal of Applied Econometrics* 8(3): 231–47.

Harvey, Andrew C. and Thomas M. Trimbur, 2003. "General Model-Based Filters for Extracting Cycles and Trends in Economic Time Series." *The Review of Economics and Statistics* 85(2): 244–55.

Hausman, Jerry A., Jason Abrevaya, and Fiona M. Scott-Morton, 1998. "Misclassification of the Dependent Variable in A Discrete-Response Setting." *Journal of Econometrics* 87(2): 239–69.

Heckman, James J, 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1): 153–61.

Hodrick, Robert J. and Edward C. Prescott, 1997. "Post-war U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking* 29(1): 1–16.

Hogue, Carma R., 1985. "History of the Problems Encountered in Estimating Gross Flows." In "Proceedings of the Conference on Gross Flows in Labor Force Statistics," Washington, DC.

Hogue, Carma R. and Paul O. Flaim, 1986. "Measuring Gross Flows in the Labor Force: An Overview of a Special Conference." *Journal of Business and Economic Statistics* 4(1): 111–21.

Holzer, Harry J. and Robert J. LaLonde, 1999. "Job Change and Job Stability among Less-Skilled Young Workers." Discussion Paper 1191-99, Institute for Research on Poverty.

Jaeger, David A., 1997. "Reconciling the Old and New Census Bureau Education Questions: Recommendations for Researchers." *Journal of Business and Economic Statistics* 12(3): 300–9.

Kaitz, Hyman B., 1970. "Analyzing the Length of Spells of Unemployment." *Monthly Labor Review* 93(11): 11–20.

Katz, Arnold, Klaus Teuter, and Philip Sidel, 1984. "Comparison of Alternative Ways of Deriving Panel Data from the Annual Demographic Files of the Current Population Survey." *Review of Public Data Use* 12(1): 35–44.

Kiefer, Nicholas M., 1988. "Economic Duration Data and Hazard Functions." *Journal of Economic Literature* 26(2): 646–79.

Kletzer, Lori G., 1998. "Job Displacement." *The Journal of Economic Perspectives* 12(1): 115–36.

Koopman, Siem Jan, Andrew C. Harvey, Jurgen A. Doornik, and Neil Shephard, 2007. *Structural Time Series Analyzer, Modeller and Predictor: STAMP 8*. London: Timberlake Consultants, Ltd.

Lansing, John B. and James N. Morgan, 1967. "The Effect of Geographical Mobility on Income." *The Journal of Human Resources* 2(4): 449–60.

Madrian, Brigitte C. and John Lars Lefgren, 2000. "An Approach to Longitudinally Matching Current Population Survey (CPS) Respondents." *Journal of Economic and Social Measurement* 26: 31Ű–62.

Moscarini, Giuseppe and Kaj Thomsson, 2008. "Occupational and Job Mobility in the US." *Scandinavian Journal of Economics* 109(4): 807–36.

Murray, Christian J., 2003. "Cyclical Properties of Baxter-King Filtered Time Series." *The Review of Economics and Statistics* 85(2): 472–6.

Nagypál, Eva, 2004. "Worker Reallocation over the Business Cycle: The Importance of Job-to-Job Transitions." Unpublished paper, Northwestern University.

Neal, Derek, 1999. "The Complexity of Job Mobility among Young Men." *Journal of Labor Economics* 17(2): 237–261.

Nekarda, Christopher J. and Garey Ramey, 2007. "Total Separations." Unpublished paper, University of California, San Diego.

Neumark, David and Daiji Kawaguchi, 2004. "Attrition Bias in Labor Economics Research using Matched CPS Files." *Journal of Economic and Social Measurement* 29(4): 445–72.

Peracchi, Franco and Finis Welch, 1994. "Trends in Labor Force Transitions of Older Men and Women." *Journal of Labor Economics* 12(2): 210–42.

Perry, George L., 1972. "Unemployment Flows in the U.S. Labor Market." *Brookings Papers on Economic Activity* 2: 245–92.

Petersen, Trond, 1991. "Time-Aggregation Bias in Continuous-Time Hazard-Rate Models." *Sociological Methodology* 21: 263–90.

Petersen, Trond and Kenneth W. Koput, 1992. "Time-Aggregation Bias in Hazard-Rate Models with Covariates." *Sociological Methods and Research* 21(1): 25–51.

Poterba, James and Lawrence H. Summers, 1984. "Response Variation in the CPS: Caveats for the Unemployment Analysis." *Monthly Labor Review* 107(3): 37–43.

Poterba, James and Lawrence H. Summers, 1986. "Reporting Errors and Labor Market Dynamics." *Econometrica* 54(6): 1,319–38.

Ramey, Garey, 2008. "Exogenous vs. Endogenous Separation." Unpublished paper, University of California, San Diego.

Ravn, Morten O. and Harald Uhlig, 2002. "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations." *The Review of Economics and Statistics* 84(2): 371–80.

Schlottmann, Alan M. and Henry W. Herzog Jr., 1981. "Employment Status and the Decision to Migrate." *The Review of Economics and Statistics* 63(4): 590–98.

Schlottmann, Alan M. and Henry W. Herzog Jr., 1984. "Career and Geographic Mobility Interactions: Implications for the Age Selectivity of Migration." *The Journal of Human Resources* 19(1): 72–86.

Shimer, Robert, 2005. "Reassessing the Ins and Outs of Unemployment." Unpublished paper, University of Chicago.

Shimer, Robert, 2007. "Reassessing the Ins and Outs of Unemployment." Unpublished paper, University of Chicago.

Topel, Robert H. and Michael P. Ward, 1992. "Job Mobility and the Careers of Young Men." *The Quarterly Journal of Economics* 107(2): 439–479.

Tucker, C. Jack and William L. Urton, 1987. "Frequency of Geographic Mobility: Findings From the National Health Interview Survey." *Demography* 24(2): 265–70.

U.S. Census Bureau, 2007. "Annual Geographical Mobility Rates, By Type of Movement: 1947–2007." Internet release date 16 October 2007. Retrieved 16 November 2008 from `http://www.census.gov/population/socdemo/migration/`.

Welch, Finis, 1993. "Matching the Current Population Surveys." *Stata Technical Bulletin* 12(2): 7–11.

Westat, 1998. "Survey of Income and Program Participation Quality Profile, Third Edition." SIPP Working Paper 230, U.S. Department of Commerce, Bureau of the Census.

Westat, 2001. *Survey of Income and Program Participation Users' Guide*. Washington, DC: U.S. Census Bureau, third edition.

Yashiv, Eran, 2007. "U.S. Labor Market Dynamics Revisited." *Scandinavian Journal of Economics* 109(4): 779–806.