# The Dual U.S. Labor Market Uncovered

Hie Joo Ahn

Federal Reserve Board

Bart Hobijn

Federal Reserve Bank of Chicago

Ayşegül Şahin

University of Texas at Austin and NBER*

Draft: October 17, 2022

**Abstract**

Aggregate U.S. labor market dynamics are well approximated by a dual labor market supplemented with a third home-production segment. We estimate a Hidden Markov Model, a machine-learning method, to uncover this structure. The different market segments are identified through (in-)equality constraints on labor market transition probabilities. This method yields time series of stocks and flows for the three segments for 1980-2021. Primary sector workers, who make up around 55 percent of the population, are almost always employed and rarely experience unemployment. The secondary sector, which constitutes 14 percent of the population absorbs most of the short-run fluctuations in the labor market, both at seasonal and business cycle frequencies. Workers in this segment experience 6 times higher turnover rates than those in the primary tier and are 10 times more likely to be unemployed than their primary counterparts. The tertiary segment consists of workers who infrequently participate in the labor market but nevertheless experience unemployment when they try to enter the labor force. While we find that young workers, racial minorities, and workers with lower educational attainment are more likely to belong to the secondary sector, the bulk of labor market segment variation across individuals cannot be explained by observables. Our findings imply that aggregate stabilization policies, such as monetary policy, predominantly work through the small but turbulent secondary market.

*JEL classification codes:* J6, J20.

*Keywords:* Dual labor markets, Hidden Markov Models, machine learning.

*This material is based upon work supported by the National Science Foundation under Grant No. SES-2048713. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or of any of the institutions that the authors are affiliated with.

# 1 Introduction

We show that U.S. labor market dynamics at the macro and individual levels are well characterized by a Dual Labor Market (DLM) supplemented with a tertiary home-production sector that consists of those who only infrequently participate. We uncover the dual labor market structure of the U.S. labor market by estimating a Hidden Markov Model (HMM) with inequality restrictions using labor market histories of *all* individuals in the Current Population Survey (CPS) for 1980-2021. Our paper is the first one that adopts this novel approach for the analysis of dualism in the U.S. labor market. This stark characterization sheds light on various puzzling features of the U.S. labor market and has distinct policy implications.

The DLM Hypothesis was first posited by Doeringer and Piore (1970), who argued that a useful characterization of the U.S. labor market is that of one segmented into a *primary* and a *secondary* tier. Jobs in the primary tier generally have low turnover, pay high wages, come with benefits, offer potential for job advancement, and provide job security. Jobs in the secondary tier have high turnover, pay low wages, come with limited benefits, offer few career opportunities, and provide little job security (Piore, 1970). After a flurry of papers about dualism in the labor market in the '70s and '80s,[1] the DLM Hypothesis fell into disfavor among macroeconomists during the Neoclassical Renaissance of the '80s and '90s. As early critics put it, theories of the DLM are "... too varied, incomplete, and amorphous "(Cain, 1975) to be captured in a set of microfounded first principles that explain the reasons for the endogenous emergence of discontinuous segments in the labor market (Wachter, 1974).

Recent analyses of dualism in the labor market in developed economies have mainly focused on Europe (Costain *et al.* , 2010; Bentolila *et al.* , 2019) and ignored dualism in the U.S.. This is because the institutional reasons for dualism in European labor markets, like size-dependent policies (Guner *et al.* , 2008), unionization (Berger *et al.* , 1980), and tiered contracts (Bentolila *et al.* , 2019), are much less applicable in the U.S.. However, as the theories by Bulow and Summers (1986), Albrecht and Vroman (1992), and Saint-Paul (1997) point out, dualism can emerge as a result of frictions, the existence of efficiency wages and, more generally, due to the nature of demand fluctuations in different segments of the economy even in the absence of such institutional arrangements and structures. While the regulatory and institutional differences between segments in European labor markets allow for a clear identification of which workers are in the primary and secondary tiers, the absence of such differences in the U.S. makes this identification much harder.[2] Moreover, the DLM hypothesis focuses specifically on those

[1] See, for example Reich *et al.* (1973), Harrison and Sum (1979), Berger *et al.* (1980), and Dickens and Lang (1985).

[2] Some authors have used occupation as a proxy for the labor market segment workers are in (e.g. McNabb

participating in the labor market. The importance of the participation margin for labor market fluctuations requires augmenting it with a tertiary home production sector.

We apply an unsupervised machine learning method that involves estimating a Hidden Markov Model (HMM) to identify which respondents in the Current Population Survey (CPS) are part of the primary, secondary, and tertiary sectors over the period 1980-2021. Our analysis builds on a small, but growing, literature that aims to identify a limited set of worker types to capture the relevant aspects of macro heterogeneity in the U.S. labor market. (Hall and Kudlyak, 2019; Gregory *et al.* , 2021; Shibata, 2019).

Our method differs from those used in these papers in four important ways. First, our hidden states have a direct economic interpretation that stems from the identifying restrictions we impose that are based on the DLM Hypothesis.[3] Second, in contrast to the models of Hall and Kudlyak (2019), Gregory *et al.* (2021) and Shibata (2019), we estimate monthly *time series* of the stocks and flows for each of the hidden states. Therefore, we can analyze seasonality, business cycle properties, and long-run trends in the three labor market segments we identify. Third, for our identification we use detailed labor force status data in the CPS to inform our hidden states.[4] Fourth, the method yields individual-level results that aggregate to the monthly labor market stocks and flows published by the Bureau of Labor Statistics (BLS).

The aggregate results show that the U.S. labor market is well characterized by three distinct tiers. Workers, in the *primary* segment, who make up around 55 percent of the population, are almost always employed and they very rarely experience unemployment. They also seamlessly move from non-participation to employment unlike workers in the secondary and tertiary sectors. Labor market frictions are basically irrelevant for these *primary* sector workers. The *secondary* sector, which constitutes 14 percent of the population, exhibits high turnover and high unemployment and absorbs most of the short-run fluctuations in the labor market, at both seasonal and business cycle frequencies. Workers in this sector are 6 times more likely to move between labor market states than those in the primary tier and are 10 times more likely to be unemployed than their primary counterparts. The *tertiary* sector includes workers who are only loosely attached to the labor market and has a very low employment-to-population ratio. These workers tend to experience unemployment when they enter the labor force from nonparticipation but do not share the high job-loss rate of secondary workers. These large differences

and Psacharopoulos, 1981).

[3] A useful analogy is the structural VAR literature. The structural VAR approach helps recover the structural estimate of parameters of interest by imposing restrictions to a reduced-form model that are informed by economic theory.

[4] We are the first to systematically introduce refined labor force states, such as part-time for economic reasons, discouraged or marginally attached, into a unified statistical framework used to identify underlying heterogeneity in the labor market.

between the three tiers of the labor market imply that average stocks and flow rates, which are commonly used to quantitatively discipline macroeconomic models of the labor market, are not at all reflective of individual labor market experiences and outcomes.

The stark differences between the three segments mean that each segment contributes to different aspects of aggregate labor market outcomes. The primary sector accounts for more than 80% of employment and participation but its contribution to the unemployment rate is much smaller. Only about a quarter of aggregate unemployment is due to the incidence of unemployment in the primary market. What is probably the most striking finding is that the secondary market, which makes up only 13.6% of total employment, accounts for 60% of unemployment in the economy. Labor market dynamism, for which we use a new measure of flows per capita, is also highly uneven, with the secondary market accounting for half of the turnover in the economy. Moreover, two of the most notable long-run labor market trends in the U.S., i.e. the trend decline in the unemployment rate and the decline in labor market dynamism, are mostly accounted by changes in the secondary sector.

We combine our estimates of individual-level posterior probabilities of being in each segment for each of the 10 million individuals in the CPS from 1980-2021 with other variables not used in our estimation. This allows us to examine the potential reasons for dualism in the U.S. labor market. We find some evidence consistent with life-cycle effects as well as discrimination. However, observable demographic characteristics only explain a small part of the cross-individual variation in segment membership. Consistent with the efficiency wage theories of dualism, analyzed in Bulow and Summers (1986), Albrecht and Vroman (1992), and Saint-Paul (1997), jobs in the primary sector are for high-skilled service occupations for which output is hard to monitor, are more stable, pay higher wages, and have higher returns to schooling and experience.

The combination of the aggregate and individual-level evidence we present in this paper points to dualism in the U.S. labor market being an equilibrium division of labor, under labor market imperfections, that minimizes adjustment costs in response to predictable seasonal as well as unpredictable business cycle fluctuations.

The rest of this paper is structured as follows. In the next section we discuss the details of our methodology in the context of the literature and explain how the DLM provides a way to think about many dimensions of micro and macro heterogeneity at the same time. Next, we describe how we distinguish the primary, secondary, and tertiary markets in the context of an HMM and how we resolve the practical challenge of estimating the model with many parameters and observations subject to the identifying restrictions we impose. We present our results in two parts. In Section 4 we show how the primary, secondary, and tertiary tiers are very different from each other as well as from the overall labor market and we quantify the importance of each of the three segments for the trends and cycles in commonly analyzed aggregates. In Section 5 analyze the individual-level evidence and discuss what it reveals about the possible causes of labor-market dualism in the U.S..

# 2 Importance and Identification of Macro Heterogeneity

The division of the population into the labor market states of employed, unemployed, and non-participants is the common classification system used to analyze macroeconomic outcomes in the labor market, including in the CPS. While these categories capture very important differences in workers' labor market experiences, they are too coarse to characterize many different aspects of individual and aggregate labor market outcomes.

A slew of recent studies has emphasized the importance of different subcategories of persons within the three labor market states for individual and aggregate outcomes. These include heterogeneity among the unemployed that accounts for duration distribution of unemployment (van den Berg and van Ours, 1996; Hornstein, 2012; Ahn and Hamilton, 2020a; Kroft *et al.* , 2016; Elsby *et al.* , 2015), heterogeneity in the type of jobs for the employed to account for the tenure distribution (Hall, 1982; Hyatt and Spletzer, 2016) as well as worker turnover (Pries, 2004; Pries and Rogerson, 2021), and heterogeneity among different categories of non-participants and unemployed to account for fluctuations in matching efficiency (Hall and Schulhofer-Wohl, 2018; Sedlacek, 2016). All these studies have the common implication that a more accurate description of individual-level labor market histories as well as macro-level labor market dynamics requires the identification and measurement of broad subcategories of the three coarse labor market states. We refer to these subcategories as "Macro Heterogeneity."

Relatedly, commonly used models with search frictions (e.g. Mortensen and Pissarides, 1994; Shimer, 2005) imply that flows between employment and unemployment are Markovian in that multi-period transition probabilities are compounded one-period transition probabilities, where the latter are calibrated from the data. As Kudlyak and Lange (2017) and Morchio (2020) point out, this is neither the case for employment-unemployment flows in the data nor for flows across the participation margin. As a result, such models do not fit individual multi-period transition probabilities between labor market states. One approach to fit these individual histories is to represent them as a mixture of different unobserved first-order Markov processes. Such a mixture approach is not only useful to match individual-level evidence, Ferraro (2018) and Gregory *et al.* (2021) show that mixtures of commonly-used models help us better understand the sources of persistence and asymmetries in aggregate labor market dynamics.[5]

These insights provide the following research challenge for the identification of Macro Heterogeneity: Develop a method to find a parsimonious representation of individual and aggregate labor market dynamics in terms of a mixture of a limited number of hidden first-order Markov processes that each have a clear economic interpretation. This method, by definition, involves classifying individuals at each point in time into untagged hidden labor market states. Because the algorithm does not use prior information about who belongs to which group, it is a form of *unsupervised* machine learning.

The method we use to tackle this challenge is a Hidden Markov Model, which is a statistical tool that estimates latent states and their dynamics from data on categorical sequences. We use this method because it has four important advantages over earlier studies of Macro Heterogeneity in the U.S. labor market (Hall and Kudlyak, 2019; Shibata, 2019; Gregory *et al.* , 2021).

The first is that it allows us to impose specific identifying restrictions, guided by the Dual Labor Market Hypothesis posited in Doeringer and Piore (1970), across the hidden states we identify that assure that they are interpretable as the primary, secondary, and tertiary sectors of the labor market. These identifying restrictions are (in-)equality constraints on the persistence of and turbulence between labor force states.

The implementation of an HMM with inequality restrictions is appealing since it mimics the use of similar restrictions for the identification of economically meaningful shocks in Structural Vector Autoregression models (SVAR) as in Stock and Watson (2001), Christiano *et al.* (2006), and Baumeister and Hamilton (2015). However, it has not been yet been applied in labor economics since its implementation is numerically challenging. Our main methodological contribution is to show that this can be done through a generalization of the Baum-Welch (BW) algorithm, introduced by Baum *et al.* (1970) and Welch (2003), that is commonly used for the estimation of HMMs. Just like the BW algorithm, our method is an application of the Expectation-Maximization (EM) algorithm (Dempster *et al.* , 1977). The difference is that in our method the M-step involves the numerical maximization of the expectation of the complete-data likelihood function subject to the identifying (in-)equality restrictions we impose. We show that this is feasible because it can be split up into a set of well-behaved convex maximization problems for which efficient numerical methods are available.

[5]The crucial insight is that a higher-order Markov process can be characterized as a mixture of first-order Markov processes. See Granger and Morris (1976) for an example of this for ARMA processes. This insight has been applied by Ferraro (2018) and Gregory *et al.* (2021). Ferraro (2018) shows that the dynamics of a mixture of Mortensen and Pissarides (1994) models can have very rich dynamics. Gregory *et al.* (2021) show the same for a mixture of Menzio and Shi (2011) models.

The second is that our method allows us to estimate monthly time-varying stocks of and transition probabilities between the hidden states. This yields a set of stocks and flows for the hidden states that is conceptually identical to those published monthly by the BLS for the labor force states of employment, unemployment, and non-participation and studied extensively in many papers (e.g Marston, 1976; Blanchard *et al.* , 1990; Barnichon and Nekarda, 2012; Elsby *et al.* , 2015). This allows us to analyze the importance of the segmentation of the labor market for seasonality, business cycle fluctuations, and long-run trends.

The third advantage is that our HMM allows us to use an extensive set of twenty nine nuanced answers in the CPS about the types of and reasons for employment, unemployment, and non-participation that respondents provide. These include part-time versus full-time employment, the reasons for unemployment, as well as the intent to and reason for not looking for a job when not participating, among others. We use the variation in labor market outcomes between different groups of individuals like those part-time employed for economic reasons versus full-time employed, those temporarily unemployed versus ones having been laid off, or marginally attached non-participants versus those not wanting a job, to enhance our assessment of the likelihood of which of the three labor market segments they are part of.

The final advantage is that, based on the reported labor market histories in the CPS, for each individual our method provides estimates of the posterior probabilities that she or he is part of each of the three respective labor market segment. To make sure these individual-level estimates aggregate to the three-state stocks and flows published by the BLS and analyzed in other studies, we assume, just like in the published data, that missing observations are random and that workers do not make any classification errors when they report whether they are employed, unemployed, or not participating in the labor market.[6] The individual-level posterior probabilities are additional variables for all CPS respondents that can be used to assess both the incidence of segment membership by demographic group as well as the impact of segment membership on labor market outcomes, like industry and occupation of employment, earnings, as well as hours worked and tenure.

# 3 Identification of a Dual Labor Market in an HMM

Though HMMs have been used in several empirical studies of U.S. labor market data,[7] the specific application to uncover the segments of the DLM using (in-)equality constraints as

[6] There is an extensive literature on such classification errors (Abowd and Zellner, 1985; Blanchard *et al.* , 1990; Feng and Hu, 2013; Elsby *et al.* , 2015; Ahn and Hamilton, 2020b) and a large degree of disagreement about their importance. This is beyond the scope of our analysis in this paper.

[7] For example, Boeschoten *et al.* (2020), and Shibata (2019).

Table 1: Hidden states in HMM

| State | Description |
|---|---|
| EP | Primary employed |
| ES | Secondary employed |
| ET | Tertiary employed |
| UPS | Primary short-term unemployed |
| UPL | Primary long-term unemployed |
| USS | Secondary short-term unemployed |
| USL | Secondary long-term unemployed |
| UTS | Tertiary short-term unemployed |
| UTL | Tertiary long-term unemployed |
| NP | Primary non-participant |
| NS | Secondary non-participant |
| NT | Tertiary non-participant |

identifying assumptions is new in this paper. To explain our contribution, we first describe the structure of the HMM we estimate. We then discuss the identifying restrictions. Finally, we describe how we deal with these restrictions in the estimation of the model and how it does not only yield parameter estimates but also posterior probabilities that the individual survey respondents in the CPS are in each of the market segments.

## 3.1 Structure of the Dual Labor Market HMM

The structure of the HMM we estimate is guided by both the aim to estimate the stocks and flows in the segments of the DLM as well as by the specific structure of the CPS data (Flood *et al.* , 2020) we use for that purpose. We describe our benchmark specification here and discuss why we chose this specification in Subsection 3.3, where we cover model selection.

Our specification consists of three labor market tiers: A primary ($P$), secondary ($S$), and tertiary ($T$). Each of these segments themselves consist of four hidden states: employed ($EM$), short-term unemployed ($UMS$), long-term unemployed ($UML$), and non-participants ($NM$). Here $M \in \{P, S, T\}$ denotes the market segment. The resulting twelve hidden states are listed in Table 1.

Our goal is to classify persons, who are categorized as either employed, unemployment, or not-in-the-labor-force, into a set of refined hidden states based on their responses to the CPS about their labor market status. In the context of the HMM these responses are called *emissions*, because they are observable signals that respondents "send" about the hidden state they are in.

The HMM consists of two layers. The first is the stochastic process that drives the evolution of the hidden state for each individual that aggregate to the flows and stocks in the labor market.

We denote the hidden labor market state of individual $i$ by $\ell_{i,t} \in L$, where $L$ is the set of twelve hidden labor market states. It follows a first-order Markov process in that the transition probabilities satisfy

$$q_{l,l',t} = P\left(\ell_{i,t} = l' \mid \ell_{i,t-1} = l; t\right) = P\left(\ell_{i,t} = l' \mid \ell_{i,t-1} = l, \cap_{k=2}^{\infty} \ell_{i,t-k} = l_{t-k}; t\right), \tag{1}$$

where $(l, l') \in L \times L$. The argument $t = 1, \ldots, T$ reflects that they vary over time. These transition probabilities are the *flow* rates between the different hidden states in our model. These flow rates determine the evolution of the stocks of individuals in the each hidden state. These stocks are the unconditional probabilities of an individual being in state $l \in L$ in month $t$. We denote them by

$$\delta_{l,t} = P\left(\ell_{i,t} = l; t\right). \tag{2}$$

The advantage of the assumption that the hidden states follow a first-order Markov process is that this makes the hidden states interpretable as states in a generalized theoretical model of the labor market in which transitions between the states follow a first-order Markov process, as in the seminal model by Mortensen and Pissarides (1994) for example. At first glance, this assumption might seem restrictive. However, because there are more hidden states than the three observed categories of employment, unemployment, and non-participation, the observed categories are a mixture of the underlying hidden states and mixtures of first-order Markov processes can have a wide range of non-Markovian properties.

The second layer of the HMM is the stochastic process that determines the information the emissions provide about the hidden state that an individual is in. We denote the emission of individual $i = 1, \ldots, n$ in month $t$ by $x_{i,t} \in X$, where $X$ is the set of possible emissions that we discuss in more detail below. The relationship between the emissions and the hidden states is known as the emission model.

The main assumption behind the emission model in an HMM is that the probability of a particular emission only depends on the current hidden state. This conditional-independence assumption yields the following expression for the emission probability

$$\omega_{x,l,t} = P\left(x_{i,t} = x \mid \ell_{i,t} = l; t\right), \text{ where } x \in X \text{ and } l \in L. \tag{3}$$

Here, the argument $t$ captures that the emission probabilities in our model vary over time.

We include in the set of emissions, $X$, information about the labor force status, i.e. employed, unemployed, or non-participant, the type of employment, the reason for unemployment, the duration of unemployment, whether or not non-participants completed a seasonal or tem-

Table 2: Observed emissions in HMM

| Emission | Description |
|---|---|
| M | Labor market state not reported in the CPS |
| EX | Employed, no other detail |
| EPE | Employed, part-time for economic reason |
| ENW | Employed, absent for other reasons |
| UTL5 | Unemployed on temporary layoff, duration $<$ 5w |
| UTL14 | Unemployed on temporary layoff, duration $<$ 14w |
| UTL26 | Unemployed on temporary layoff, duration $<$ 26w |
| UTLLT | Unemployed on temporary layoff, duration $>$ 26w |
| UTJ5 | Unemployed temporary job ended, duration $<$ 5w |
| UTJ14 | Unemployed temporary job ended, duration $<$ 14w |
| UTJ26 | Unemployed temporary job ended, duration $<$ 26w |
| UTJLT | Unemployed temporary job ended, duration $>$ 26w |
| UJL5 | Unemployed job loser, duration $<$ 5w |
| UJL14 | Unemployed job loser, duration $<$ 14w |
| UJL26 | Unemployed job loser, duration $<$ 26w |
| UJLLT | Unemployed job loser, duration $>$ 26w |
| UX5 | Unemployed n.e.c., duration $<$ 5w |
| UX14 | Unemployed n.e.c., duration $<$ 14w |
| UX26 | Unemployed n.e.c., duration $<$ 26w |
| UXLT | Unemployed n.e.c., duration $>$ 26w |
| NTJDW | Non-participant who ended temporary job and discouraged worker |
| NTJMA | Non-participant who ended temporary job, not discouraged but marginally attached |
| NTJNA | Non-participant who ended temporary job, recently searched but not available for work |
| NTJNS | Non-participant who ended temporary job, no previous job search but want a job |
| NTJDNW | Non-participant who ended temporary job, does not want a job |
| NDW | Non-participant and discouraged worker |
| NMA | Non-participant, not discouraged but marginally attached |
| NNA | Non-participant, recently searched but not available for work |
| NNS | Non-participant, no previous job search but want a job |
| NDNW | Non-participant, does not want a job |

porary job, and information about labor-force attachment. This results in 29 different possible emissions, listed in Table 2.

The emissions distinguish between unemployed of different durations. This might seem like a violation of the conditional-independence assumption because to report having been unemployed for several months seems to imply that one was unemployed in the previous month. This, however, is not the case. Unemployed respondents in the CPS report how long they have been searching for a job rather than the duration of their unemployment spell. Many respondents in the survey report to be employed or out of the labor force during the period for which they later report to have been searching for a job (Elsby *et al.* , 2011).

In addition to the emissions listed in Table 2, there are also missing observations. The 4-8-4 panel structure of the CPS is such that for each individual $i$ who enters the sample in period $t_i$ we have observations for $t = t_i, \ldots, t_i + 15$. At least 8 of these observations, and possibly more, are missing.

To summarize, we have a panel of incomplete observed 16-month long labor market histories across individuals that sends an imperfect signal about in which of the 12 hidden labor market states they are in at each point in time. We use the HMM described above to estimate the following time series: (*i*) the share of individuals in each of the states, $\delta_{j,t}$, i.e. the equivalent of the stocks, (*ii*) the transition probabilities between the latent states $q_{l',l,t}$, i.e. the flow rates, and (*iii*) the emission probabilities, $\omega_{x,l,t}$. We denote the vector with all these parameters as $\boldsymbol{\theta}$, the vector with the observed history of emissions for individual $i$ as $\boldsymbol{x}_i$, and the vector with the unobserved path of underlying hidden states as $\boldsymbol{\ell}_i$.

## 3.2 Identification

Without any further restrictions on the parameters the model would be observationally equivalent for any permutation of the hidden states in the set $L$. This is known as "label swapping" (Allman *et al.* , 2009).[8] To avoid this and to make each element of $L$ correspond to a particular economically meaningful hidden state, we introduce three types of restrictions on the parameters of the HMM we estimate.

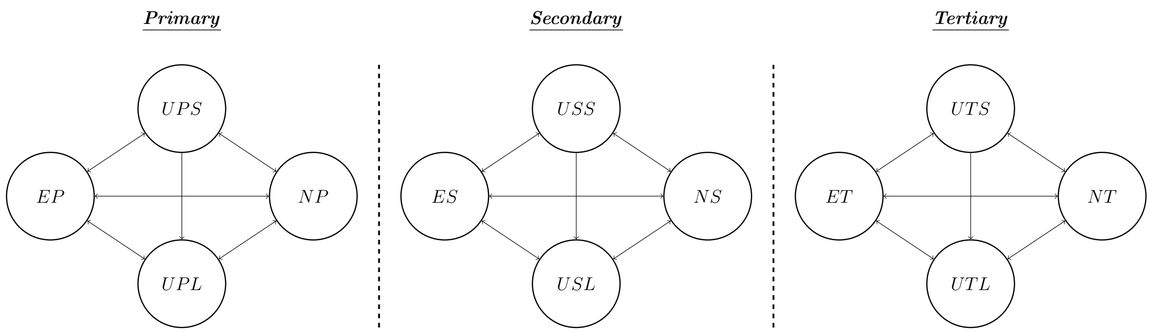
#### Inequality restrictions on transition probabilities

The first type of restrictions captures the differences in relative turnover rates between labor market segments from the DLM Hypothesis.

The hypothesis is that the primary tier of the labor market is characterized by a higher level of employment stability than the secondary and tertiary tiers. Employment stability is a key attribute that distinguishes the primary market from the secondary market (Doeringer and Piore, 1970; Piore, 1970; Berger *et al.* , 1980; Dickens and Lang, 1985). According to Wachter (1974), one of the hypotheses defining the dual labor market is that workers in the secondary sector develop a pattern of job instability. Similarly, Bentolila *et al.* (2019) mention that the main feature of dual labor market in Europe is the coexistence of open-ended and fixed-term contracts. The former guarantee job security, while the latter make a job last only for a short period of time. The inequality restriction we impose captures the gist of this aspect of the DLM

[8]One approach, taken by Shibata (2019), is to estimate the unrestricted model and then label the hidden states ex-post based on the unrestricted parameter estimates. This, however, is not practical in our case with time-varying parameters, since the meaning of each of the states can change from period to period in that case, i.e. the unrestricted version of our model suffers from "temporal label swapping". A similar approach is adopted in the literature on regime-switching models (e.g., Schorfheide (2005), and Amisano and Tristani (2019)).

Figure 1: Description of the three labor market segments.




theory. To have our parameter estimates satisfy this property, we impose the restrictions that

$$q_{EP,EP,t} \geq q_{ES,ES,t} + 0.05 \text{ and } q_{EP,EP,t} \geq q_{ET,ET,t} + 0.05, \text{ for all } t. \tag{4}$$

The tertiary market is the one in which people go through persistent spells of non-participation. The previous literature on dual labor market focuses only on transitions between employment and nonemployment. Considering substantial heterogeneity in labor supply elasticities among nonemployed individuals (Veracierto, 2008; Krusell *et al.* , 2017), we impose additional inequality restrictions to capture the persistence of non-participation in the tertiary market segment. These restrictions take the form

$$q_{NT,NT,t} \geq q_{NP,NP,t} + 0.05 \text{ and } q_{NT,NT,t} \geq q_{NS,NS,t} + 0.05, \text{ for all } t. \tag{5}$$

In addition, our model specification includes more than one hidden type of unemployment in each market. To assure that the interpretation of the hidden unemployment states we uncover matches their labels, we assume that long-term unemployment is more persistent than short-term unemployment. That is

$$q_{UML,UML,t} \geq q_{UMS,UMS,t} + 0.05 \text{ where } M \in \{P, S, T\}, \text{ for all } t. \tag{6}$$

In addition, we impose that persons can only flow from short- to long-term employment and not vice-versa, i.e.

$$q_{UML,UMS,t} = 0 \text{ where } M \in \{P, S, T\}, \text{ for all } t. \tag{7}$$

The above four constraints assure us that each of the hidden states we identify has a clear economic interpretation in the context of the DLM with a tertiary home production sector. Figure 1 summarizes the market structure we estimate using HMM.

**Zero restrictions on transition probabilities**

The second type of restrictions is guided by another assumption in the DLM Hypothesis. Namely, that there is very limited mobility between labor market segments. Because of the very short histories reported in the CPS we approximate this assumption by the restriction that respondents do not switch market tiers during the 16-month period they are in the sample. This translates into a set of zero restrictions on the transition probabilities that capture that there are no flows between the primary, secondary, and tertiary markets.

**No classification errors: Zero restrictions on emission probabilities**

The third type of restriction is that we assume there are no classification errors. This is because our goal is to uncover the stocks and flows in the segments of the DLM with a tertiary home production sector that are consistent with aggregate stocks and flows published by the BLS. With that in mind, we follow the BLS and assume that respondents correctly report their labor market status of employment, unemployment, and non-participation.

If respondents always correctly report their labor market status (employed, unemployed, non-participant), then this implies that the probability that their emission does not correspond to their hidden labor market status is zero. We impose these zero restrictions on the emission probabilities for all months in our sample.

**Random missing values**

In addition to the (in-)equality restrictions, we also impose that missing values for the emissions, $x_{i,t}$, are random. That is, the probability that a respondent does not report any emissions does not depend on the hidden state she or he is in. This way of treating the missing values means that no information is gleaned from whether an observation is missing or not.[9] This assumption is, by definition, true for the 8-month reporting gap in the CPS during which respondents drop out of the sample. So, during that gap the probability of a missing emission is one no matter what the hidden state.

## 3.3 Estimation

Because the estimation involves a large number of observations, $n$, and parameters, $\dim(\boldsymbol{\theta})$, direct maximization of the likelihood function is not feasible. However, it can be accomplished

[9] Alternatively, one can treat missing observations as being in a fourth observable state and include it in the model through the emission probabilities. This is how Ahn and Hamilton (2020b) treat missing values in their analysis of measurement error in the CPS. Though they do not use an explicit HMM.

through the application of the BW algorithm (Baum *et al.* , 1970; Welch, 2003), commonly used in machine learning and estimation of HMMs. This is a specific case of the EM algorithm (Dempster *et al.* , 1977). The particular form of the algorithm we use exploits the panel data structure (Maruotti, 2011) of the CPS and takes into account the identifying (in-)equality restrictions on the parameters.

The likelihood function, $\mathcal{L}(\boldsymbol{\theta})$, we maximize is the joint probability of observing the paths, $\{\boldsymbol{x}_i\}_{i=1}^n$, for a given vector of model parameters

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} P(\boldsymbol{x}_i;\boldsymbol{\theta})^{w_i} = \prod_{i=1}^{n} \left[ \sum_{\ell_{i,t_i+15}\in L} P(\boldsymbol{x}_i \cap \ell_{i,t_i+15};\boldsymbol{\theta}) \right]^{w_i} = \prod_{i=1}^{n} \left[ \sum_{\ell\in L} \alpha_{i,15}(\ell;\boldsymbol{\theta}) \right]^{w_i} \tag{8}$$

Here $w_i$ is the sample weight for individual $i$.[10]

$$\alpha_{i,k}(\ell;\boldsymbol{\theta}) = P(x_{i,t_i},\ldots,x_{i,t_i+k} \cap \ell_{i,t_i+k} = \ell). \tag{9}$$

It is the joint probability of the observed data from $t_i$ through $t_i + k$ *and* individual $i$ being in the latent state $\ell \in L$ at $t = t_i + k$.

In principle, the computation of $\alpha_{i,k}(\ell;\boldsymbol{\theta})$ requires the summation over all possible paths of the latent state between $t_i$ and $t_i + k$, which quickly becomes infeasible. However, the BW algorithm uses that $\alpha_{i,k}(\ell;\boldsymbol{\theta})$ can be calculated using a forward recursion. For the specific case of the CPS data with missing values, this recursion is of the form

$$\alpha_{i,0}(l) = \delta_{l,t}\left((1-\eta_{i,t_i}) + \eta_{i,t_i}\omega_{x_{i,t_i},l,t_i}\right), \text{ and} \tag{10}$$

$$\alpha_{i,k}(l') = \sum_{l\in L} \alpha_{i,k-1}(l)\, q_{l,l',t_i+k}\left((1-\eta_{i,t_i+k}) + \eta_{i,t_i+k}\omega_{x_{i,t_i+k},l',t_i+k}\right) \tag{11}$$

Here, $\eta_{i,t}$ is the indicator function for non-missing observations. It makes sure missing observations are integrated out of the fitted path, which is consistent with the assumption that they are random.

As with any application of the EM algorithm, it involves iteratively updating the parameters to monotonically increase the likelihood function. Each iteration involves two steps. An E-step and an M-step. These steps for the estimation of a panel-data HMM have been described in

[10] Because an individual appears in the likelihood for her/his whole 16 periods labor market history, no matter whether observations are missing or not, $w_i$ is, in principle, the sampling weight of individuals conditional on them reporting their labor market state for at least one out of eight interviews. However, such a weight is not provided for the CPS data. Therefore, we approximate it by their average cross-sectional weight across all the 8 months in sample. That is, $w_i$ is the average number of persons the individual represents across the 8 rotations in which they are interviewed.

Maruotti (2011) and Shibata (2019). For this reason, we leave the details for Appendix A.

Here, we focus on two specific aspects we use in the rest of our analysis: ($i$) how the E-step provides estimates of the posterior probabilities that each of the respondents in the CPS in a particular segment of the labor market, and ($ii$) how we implement the identifying (in-)equality restrictions on the parameters in the M-step.

The starting point for the EM algorithm is the complete-data log-likelihood function, which is the log of the likelihood function for the case in which all data, i.e. $\{\boldsymbol{x}_i, \boldsymbol{\ell}_i\}_{i=1}^n$, are observed. If we had data on the hidden state, we could construct the dummy variables

$$u_{i,t,l} = \mathbb{1}\left(\ell_{i,t} = l\right) \text{ and } v_{i,t,l,l'} = \mathbb{1}\left(\ell_{i,t-1} = l \cap \ell_{i,t} = l'\right). \tag{12}$$

Given these indicator functions, the complete-data log-likelihood function equals

$$\begin{aligned} \ln \mathcal{L} \;=\; & \sum_{i=1}^{n} w_i \left\{ \sum_{l \in L} u_{i,t_i,l} \ln \delta_{l,t_i} + \sum_{k=1}^{15} \sum_{l' \in L} \sum_{l \in L} v_{i,t_i+k,l,l'} \ln q_{t_i+k,l,l'} \right. \\ & \left. + \sum_{k=0}^{15} \eta_{i,t_i+k} \sum_{l \in L} u_{i,t_i+k,l} \ln \omega_{x_{i,t_i+k},l,t_i+k} \right\}. \end{aligned} \tag{13}$$

**Individual-level posterior probabilities from E-step**

In the E-step, the expectation of the complete-data log-likelihood conditional on the observed data $\boldsymbol{x} = \{\boldsymbol{x}_i\}_{i=1}^n$ and parameter vector $\boldsymbol{\theta}$ is calculated.

Taking the conditional expectation of (13) involves replacing $u_{i,t_i+k,l}$ and $v_{i,t_i+k,l,l'}$ with their conditional expectations, which we denote by $\hat{u}_{i,t_i+k,l}$ and $\hat{v}_{i,t_i+k,l,l'}$ respectively. They are calculated using the Forward-Backward recursions, part of the BW algorithm, described in Appendix A.

For our analysis it is important to realize that these conditional expectations are not only useful for the implementation of the BW algorithm. They also allow us to do individual-level analyses of our results.

The reason is that $\hat{u}_{i,t_i+k,l}$ can be interpreted as the posterior probability that a person is in a particular hidden state at time $t_i + k$, i.e.

$$\hat{u}_{i,t_i+k,l} = E\left[\mathbb{1}\left(\ell_{i,t_i+k} = l\right) \mid \boldsymbol{x}_i, \boldsymbol{\theta}\right] = P\left(\ell_{i,t_i+k} = l \mid \boldsymbol{x}_i, \boldsymbol{\theta}\right) \text{ for } l \in L. \tag{14}$$

Thus, the BW algorithm does not only yield a set of parameter estimates. For these estimates it also provides posterior probabilities of the stocks for each of the individuals in the data.

Note that the algorithm does not classify individuals in a particular hidden state at each point in time. Instead, their classification is a probabilistic assessment based on the limited information revealed by a person's labor market history from the 4-8-4 survey structure of the CPS.

Our focus, in particular, is on the posterior probability that a respondent is part of one of the three market segments. This probability is given by

$$P_i(M) = \sum_{l \in \{EM, UMS, UML, NM\}} P(\ell_{i,t} = l \mid \boldsymbol{x}_i, \boldsymbol{\theta}), \text{ where } M \in \{P, S, T\}. \quad (15)$$

Because we impose the restriction that individuals cannot flow from one market segment to another, this probability is constant over time.[11]

Our estimation procedure thus yields two additional variables for each respondent in the CPS that reflect the posterior probabilities that she or he is part of the primary or secondary segment of the labor market.[12]

### Imposing identifying zero- and inequality restrictions in M-step

The use of zero- and inequality constraints on the transition and emission probabilities is at the heart of our identification strategy to provide specific economic meaning to the hidden states we uncover. In the M-step the expectation of the complete-data likelihood function is maximized with respect to the parameters subject to these restrictions. In the absence of these restrictions, the M-step yields a well-known closed-form solution that is easy to solve, even in the case of a very large number of parameters (e.g. Maruotti, 2011). However, this is not the case under the constraints that we impose. Zero restrictions on the transition and emission probabilities are easily imposed in the maximization problem. The challenge is how to deal with the inequality constraints, especially in light of the large number of parameters we estimate.

One approach of dealing with inequality constraints in the BW algorithm is to transform the problem to one that has a closed-form solution (e.g. Levinson *et al.* , 1983; Otterpohl, 2002). This, however, is not feasible for the large number of parameters and restrictions in our model specification. Instead, we use that the maximization problem in the M-step can be split up into $3T$ sub-problems. Each of these involves the calculation of a Weighted Analytic Center and can be easily solved using the numerical method introduced in Andersen *et al.* (2011).[13]

Thus, even though our specification has a large number of parameters and constraints, we are able to impose the identifying restrictions in the M-step by reframing the maximization

[11] See Appendix A for a proof.

[12] The probability that the respondent is in the tertiary market is implied by the first two by the constraint that the probabilities add up to one.

[13] We discuss the details of this approach in Appendix A.

problem as a set of much smaller, well behaved, maximization problems for particular subsets of the parameters.

### Model selection

We estimate our model for all respondents, $i = 1, \ldots, n$, in the CPS from 1980-2021. The resulting sample size is $n = 10,178,593$ individual 4-8-4 labor market histories. Our benchmark specification has two main characteristics: $(i)$ It has three labor-market segments, and $(ii)$ each segment has two types of unemployed persons. Our choice of this model as our benchmark is based on a comparison of the model with several alternative specifications, all estimated using the 29 emissions listed in Table 2.

Table 3 lists the relevant statistics for the benchmark model and the three most notable alternative models we compared it to.[14] The table lists the number of labor market segments, hidden states, parameters, the resulting log-likelihood value, as well as two information criteria for each model. Models with lower values for the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are preferred over those with higher ones.

The top line contains the statistics for the benchmark model for which we present our estimates in the rest of this section. As discussed in the previous section, it consists of three labor market tiers with 12 hidden states. The total number of parameters is 90,216, which is 179 for each of the 504 months in the sample.

The second line in the table provides a baseline for comparison. It is the First-Order Markov (FOM) model in which there is one labor market with the three observed states of employment, unemployment, and non-participation for which transitions between these states are assumed to follow a first-order Markov process.[15] The information criteria show that this baseline is clearly rejected compared to our benchmark of the DLM with tertiary sector. This is consistent with the evidence in Kudlyak and Lange (2017) who point out that observed individual-level transitions between $E$, $U$, and $N$ are not first-order Markovian. A property that the benchmark model can capture but the FOM baseline, by assumption, cannot.

The third line shows the results for a specification of a "pure" DLM model that does not have a tertiary home production sector. This 2-segment specification lumps the primary and tertiary sectors together because both of them both having low turnover rates. Both the AIC and BIC indicate that the 2-segment specification is not preferred over the benchmark.

[14] In principle, there are many different permutations of specifications possible. We focus on the most relevant here.

[15] For example, in the Mortensen and Pissarides (1994) model transitions between employment and unemployment follow a first-order Markov process. See Shibata (2019) for a further discussion of the usefulness of the FOM model as a baseline.

Table 3: Comparison of model specifications

| | segments | states | pars | logL | AIC | BIC |
|---|---|---|---|---|---|---|
| Dual Labor Market (benchmark) | 3 | 12 | 90216 | -3.41 | 69.56 | 70.83 |
| First-Order Markov (FOM) | 1 | 3 | 18648 | -3.80 | 77.47 | 77.73 |
| Dual Labor Market without tertiary sector | 2 | 8 | 59976 | -3.46 | 70.60 | 71.45 |
| DLM, only two types of U in secondary | 3 | 10 | 68040 | -3.43 | 69.95 | 70.92 |
| DLM (benchmark) with constant transition- and emission probabilities | 3 | 12 | 5847 | -3.44 | 70.04 | 70.12 |

Source: Current Population Survey and authors' calculations.
Notes: Total number of observations is 10,178,593 CPS respondents from 1980-2021. Column definitions: *segments*: Number of labor market segments. *states*: Number of hidden states. *pars*: Number of parameters *LogL*: Mean log-likelihood across all individuals in sample. *AIC*: Akaike Information Criterion divided by 1000000. *BIC*: Bayesian Information Criterion divided by 1000000.

The fourth row of the table illustrates why we include multiple types of unemployed persons in the specification. It shows that the model with only two types of unemployed persons in the secondary tier is rejected compared to our benchmark model. The improved fit for multiple types of unemployed persons is consistent with a large number of studies that emphasize that they are necessary to match the unemployment duration distribution and the existence of long-term unemployment.[16] Other permutations of the model without multiple types of unemployed in different segments, not reported in the table, all yield higher information criteria than our benchmark specification.

The last row of the table reports the statistics from the benchmark model with constant transition and emission probabilities (constant-probability model). As we shut down time-variations in the probabilities the number of parameters reduces substantially relative to our benchmark. The AIC prefers the benchmark model over the constant probability model, while the BIC prefers the constant probability model over the benchmark model. It is because the BIC penalizes the increased number of parameters more than the AIC does. So, based on the information criteria, it is hard to distinguish between the benchmark model and the one with constant transition and emission probabilities.

The reason we use the specification with time-varying parameters as our benchmark is that it fits one-period transition probabilities in the data well, while the model with constant parameters fails to do so. The model with time varying parameters manages to fit the high-frequency variations in observed flow rates in the data by allowing the transition probabilities to fluctuate over time. Therefore, it captures seasonal and business cycle fluctuations in the flow rates very well while the model with constant parameters smoothes out these fluctuations.[17]

**Reliability of classification of individuals across segments**

Though the likelihood-based model selection criteria in Table 3 provide useful information about which model better fits the data, they are not meant to capture how reliably the model classifies individuals in the market segments it identifies. For this purpose, we introduce an additional summary statistic.

This statistics is based on the distance between the estimated the posterior distribution across labor market segments for each individual and the uniform distribution. To see why this distance captures the reliability of the classification of an individual, it is useful to write the

[16]For example, Hornstein (2012), Kroft *et al.* (2016), and Ahn and Hamilton (2020a).

[17]See, for example, Figure B.1 in Appendix B the unemployment-to-employment transition rates for these two specifications.

estimated posterior distribution across the three markets as the triple

$$\{P_i(P), P_i(S), P_i(T)\}. \tag{16}$$

If the data provide no information on which of the segments an individual is in then this tuple is the uniform distribution $\{1/3, 1/3, 1/3\}$. If, on the other hand, the data perfectly pin down the tier then this triple is either $\{1,0,0\}$, $\{0,1,0\}$, or $\{0,0,1\}$, depending on whether the individual is in the primary, secondary, or tertiary market respectively.

To measure the degree of information the model provides about the segment membership of individual $i$ in the sample, we calculate the rescaled distance of the posterior distributions from the non-informative, uniform, case. This measure is given by

$$D_i = \frac{\sqrt{9}}{\sqrt{6}} \sqrt{\sum_{M \in \{P,S,T\}} \left(P_i(M) - 1/3\right)^2} \in [0,1]. \tag{17}$$

It is zero if the model does not provide any information about the segment membership of individual $i$ and one if it is fully informative. Figure 2 shows the distribution of $D_i$ across all CPS respondents for our baseline model. It shows that $D_i \geq 0.99$ for more than 40 percent of them and $D_i > 0.95$ for more than half of the respondents. Thus, the model is able to reliably classify the bulk of the individuals in different segments.[18]

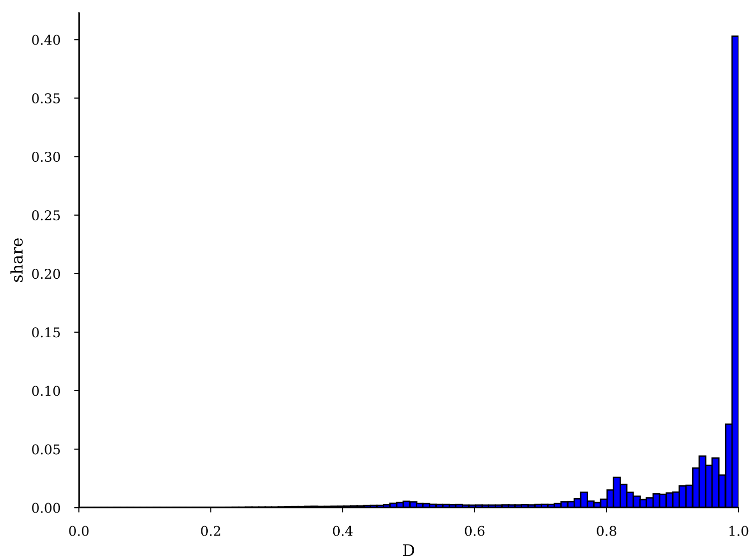



Figure 2: Distribution of $D_i$ across CPS respondents

*Source:* CPS and authors' calculations.

[18] The distribution of the three posterior probabilities is shown in Figure B.2 in the Appendix.

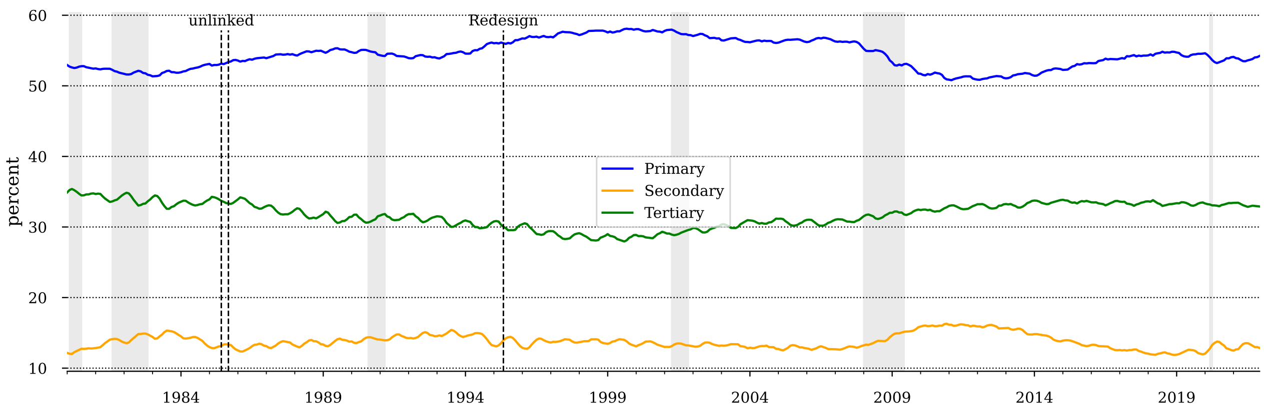
To illustrate the information that the algorithm distills from the detailed 29 emissions we use, consider Table 4. It contains four examples of how the probabilities that the model assigns to CPS respondents being part of the three market segments evolve depending on their reported history of emissions. All examples are for respondents in the sample from January 2005 through April 2006. Example I is for someone who reports to be employed, not part-time for economic reasons and not absent from work, for all eight months she is in the survey. In the first month in the sample the model assigns an 89.2 percent probability this person is in the primary market and 7.4 percent and 3.4 for the secondary and tertiary markets respectively. Because of a lack of history in the first month, these are the unconditional probabilities of someone who reports this emission being in each of the market segments. As the individual continues to report the same type of employment for the subsequent months the likelihood that she is in the primary market increases for two reasons. The first is that longer employment spells are more likely in the primary market. The second is that those in the primary segment are more likely to report they are not absent from work and do not work part-time for economic reasons. The combination of these two effects yields a posterior probability, based on the whole reported history, of 99.7 percent that the person is in the primary market.

To illustrate the information that the extended 29 emissions provide in addition to the three basic labor-force statuses of employed, unemployed, and non-participation, consider Example II. Just like in Example I, the respondent reports to be employed in all eight months that they are in the sample. However, in the middle six months they report to be part-time employed for economic reasons. Even though this respondent is always employed, the algorithm assigns them with almost certainty to the secondary segment after their eight months in the sample. This is because others who report to be part-time employed for economic reasons are likely to flow into unemployment and non-participation and are therefore more likely to be in the secondary tier.

Thus, we get substantial identification out of the fact that our estimates imply very different probabilities of people reporting different emissions depending on what market segment they are in. This can be seen from Table 5. It reports the average emission probabilities, $\bar{\omega}_{x,l}$, over our sample period. For this specific example, what matters is that, on average, only 1 percent of those employed in the primary sector report to be so part-time for economic reasons (element (EPE,EP) in the table) while about a third of those employed in the secondary segment say they are ((EPE,ES) in the table). The benefit of using 29 emissions, rather than 3, in terms of the reliability with which CPS respondents are classified is reflected in $\bar{D}$. It is 0.914 for our benchmark model with 29 emissions and 0.895 for the same specification using only 3 of them.

Example III shows that those persistently are nonparticipants who do not want a job are

Figure 3: Share of population in each labor market segment.




classified in the tertiary sector.

As we showed in Figure 2, not all respondents get classified with the level of reliability of those in the first three examples. Example IV shows a mixed employment, unemployment, and non-participation history that results in more uncertainty about which segment the respondent belongs to. The advantage of our method is that it does not force us to classify this person in a particular segment. Instead, for both the aggregate- and individual-level results we present in the next sections, we simply weight respondents by their posterior probabilities across the three segments. Thus, we use the estimated individual-level posterior probabilities to construct aggregates for each of the market segments and analyze how they differ in terms of the levels of their unemployment and participation rates, turnover, as well as their contributions to overall labor market outcomes.

# 4 Characteristics of each of the market segments

The most important finding from our analysis is that the U.S. labor market can be thought of as being comprised of three distinct segments, each of which is very different from the aggregate. The stark differences between the market tiers are clear from their average outcomes over time.[19] These averages are listed in Table 6, which provides them for the main labor market aggregates for each segment as well as for the labor market as a whole. Interestingly, as Figure B.5a shows, there is little variation on the share of population in three labor market segments over time.

The last row of the table reports a new labor market aggregate that we introduce to measure

[19] We focus on these averages here and provide the underlying estimated time series in Figures B.3 through B.6 in Appendix B.

Table 4: Inference of market segments based on emissions

| Date | Emission | P(P) | P(S) | P(T) |
|---|---|---|---|---|
| | Example I | | | |
| 2005-01 | Employed-not PTER+no other absence | 89.2 | 7.4 | 3.4 |
| 2005-02 | Employed-not PTER+no other absence | 92.6 | 4.9 | 2.5 |
| 2005-03 | Employed-not PTER+no other absence | 94.8 | 3.3 | 1.9 |
| 2005-04 | Employed-not PTER+no other absence | 96.4 | 2.2 | 1.4 |
| 2006-01 | Employed-not PTER+no other absence | 98.9 | 0.9 | 0.2 |
| 2006-02 | Employed-not PTER+no other absence | 99.3 | 0.6 | 0.1 |
| 2006-03 | Employed-not PTER+no other absence | 99.5 | 0.4 | 0.1 |
| 2006-04 | Employed-not PTER+no other absence | 99.7 | 0.3 | 0.1 |
| | Example II | | | |
| 2005-01 | Employed-not PTER+no other absence | 89.2 | 7.4 | 3.4 |
| 2005-02 | Employed-PTER | 31.5 | 66.3 | 2.2 |
| 2005-03 | Employed-PTER | 1.7 | 98.2 | 0.1 |
| 2005-04 | Employed-PTER | 0.1 | 99.9 | 0.0 |
| 2006-01 | Employed-PTER | 0.0 | 100.0 | 0.0 |
| 2006-02 | Employed-PTER | 0.0 | 100.0 | 0.0 |
| 2006-03 | Employed-PTER | 0.0 | 100.0 | 0.0 |
| 2006-04 | Employed-not PTER+no other absence | 0.0 | 100.0 | 0.0 |
| | Example III | | | |
| 2005-01 | Nonparticipants who do not want a job | 4.4 | 7.2 | 88.4 |
| 2005-02 | Nonparticipants who do not want a job | 2.3 | 3.2 | 94.5 |
| 2005-03 | Nonparticipants who do not want a job | 1.1 | 1.5 | 97.4 |
| 2005-04 | Nonparticipants who do not want a job | 0.5 | 0.7 | 98.8 |
| 2006-01 | Nonparticipants who do not want a job | 0.0 | 0.1 | 99.8 |
| 2006-02 | Nonparticipants who do not want a job | 0.0 | 0.1 | 99.9 |
| 2006-03 | Nonparticipants who do not want a job | 0.0 | 0.0 | 100.0 |
| 2006-04 | Nonparticipants who do not want a job | 0.0 | 0.0 | 100.0 |
| | Example IV | | | |
| 2005-01 | Employed-not PTER+no other absence | 89.2 | 7.4 | 3.4 |
| 2005-02 | U-Temporary job ended-less than 5 weeks | 60.3 | 39.6 | 0.0 |
| 2005-03 | Nonparticipants who do not want a job | 45.1 | 54.8 | 0.0 |
| 2005-04 | Nonparticipants who do not want a job | 45.9 | 54.0 | 0.1 |
| 2006-01 | Nonparticipants who do not want a job | 10.4 | 88.7 | 0.8 |
| 2006-02 | Nonparticipants who do not want a job | 11.2 | 86.9 | 2.0 |
| 2006-03 | Nonparticipants who do not want a job | 12.0 | 83.5 | 4.5 |
| 2006-04 | Nonparticipants who do not want a job | 11.2 | 79.6 | 9.2 |

Source: Current Population Survey and authors' calculations.
Notes: Imputed probabilities of being in primary, secondary, or tertiary market segment for hypothetical emissions history.

Table 5: Average emission probabilities

| State<br>Emission | EP | ES | ET | UPS | UPL | USS | USL | UTS | UTL | NP | NS | NT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EX | 98.7 | 68.1 | 93.4 | - | - | - | - | - | - | - | - | - |
| EPE | 1.0 | 29.4 | 2.2 | - | - | - | - | - | - | - | - | - |
| ENW | 0.3 | 2.5 | 4.4 | - | - | - | - | - | - | - | - | - |
| UTL5 | - | - | - | 29.4 | 0.6 | 8.0 | 0.2 | 4.0 | 0.3 | - | - | - |
| UTL14 | - | - | - | 4.6 | 13.3 | 4.1 | 0.4 | 0.7 | 1.0 | - | - | - |
| UTL26 | - | - | - | 0.4 | 6.6 | 1.0 | 0.6 | 0.4 | 0.4 | - | - | - |
| UTLLT | - | - | - | 0.4 | 3.8 | 0.8 | 1.7 | 0.5 | 0.2 | - | - | - |
| UTJ5 | - | - | - | 9.3 | 0.3 | 5.4 | 0.2 | 0.9 | 0.1 | - | - | - |
| UTJ14 | - | - | - | 0.7 | 3.2 | 2.7 | 1.9 | 0.3 | 0.3 | - | - | - |
| UTJ26 | - | - | - | 0.2 | 1.4 | 0.6 | 1.7 | 0.1 | 0.1 | - | - | - |
| UTJLT | - | - | - | 0.4 | 0.3 | 0.2 | 5.0 | 0.5 | 0.1 | - | - | - |
| UJL5 | - | - | - | 32.0 | 2.9 | 9.6 | 1.2 | 1.5 | 0.2 | - | - | - |
| UJL14 | - | - | - | 1.5 | 33.3 | 6.8 | 4.0 | 0.8 | 0.4 | - | - | - |
| UJL26 | - | - | - | 0.2 | 19.7 | 1.3 | 6.0 | 0.4 | 0.2 | - | - | - |
| UJLLT | - | - | - | 0.9 | 11.6 | 0.5 | 28.6 | 1.7 | 0.3 | - | - | - |
| UX5 | - | - | - | 16.3 | 0.5 | 34.1 | 1.5 | 71.3 | 6.9 | - | - | - |
| UX14 | - | - | - | 2.1 | 1.7 | 18.6 | 9.7 | 7.7 | 54.3 | - | - | - |
| UX26 | - | - | - | 0.5 | 0.5 | 3.8 | 7.4 | 0.7 | 23.9 | - | - | - |
| UXLT | - | - | - | 1.2 | 0.2 | 2.7 | 30.0 | 8.3 | 11.4 | - | - | - |
| NTJDW | - | - | - | - | - | - | - | - | - | 0.0 | 0.1 | 0.0 |
| NTJMA | - | - | - | - | - | - | - | - | - | 0.0 | 0.1 | 0.0 |
| NTJNA | - | - | - | - | - | - | - | - | - | 0.0 | 0.0 | 0.0 |
| NTJNS | - | - | - | - | - | - | - | - | - | 0.3 | 0.6 | 0.0 |
| NTJDNW | - | - | - | - | - | - | - | - | - | 1.1 | 0.9 | 0.2 |
| NDW | - | - | - | - | - | - | - | - | - | 0.9 | 3.6 | 0.1 |
| NMA | - | - | - | - | - | - | - | - | - | 1.4 | 6.9 | 0.1 |
| NNA | - | - | - | - | - | - | - | - | - | 0.4 | 1.9 | 0.1 |
| NNS | - | - | - | - | - | - | - | - | - | 10.0 | 19.0 | 1.3 |
| NDNW | - | - | - | - | - | - | - | - | - | 85.9 | 66.9 | 98.3 |

Notes: - Average probability of observed emission conditional on being in state over sample. No-classification-error restrictions are indicated by '-'.

Table 6: Labor market aggregates by segment

| | Primary | Secondary | Tertiary | Total |
|---|---|---|---|---|
| Share of population | 54.46 | 13.75 | 31.79 | 100.00 |
| Unemployment rate | 2.07 | 26.45 | 19.92 | 6.62 |
| Labor-force participation rate | 97.16 | 72.92 | 8.84 | 65.77 |
| Employment-to-population ratio | 95.15 | 53.55 | 7.05 | 61.42 |
| Flows per capita | 0.50 | 3.20 | 0.62 | 0.91 |

Source: Current Population Survey and authors' calculations.
Notes: Average of reported statistics over sample period for each market segment and total civilian non-institutionalized population 16-years and over. Flows per capita are annual flows between E,U, and N per person.

Table 7: 1- and 12-month transition probabilities in different market segments

| segment | to<br>freq<br>from | E<br>1-m | US<br>1-m | UL<br>1-m | N<br>1-m | E<br>12-m | US<br>12-m | UL<br>12-m | N<br>12-m |
|---|---|---|---|---|---|---|---|---|---|
| Primary | E | 97.92 | 0.73 | 0.03 | 1.32 | 95.08 | 0.82 | 1.21 | 2.88 |
| | US | 51.35 | 8.39 | 33.19 | 7.06 | 93.89 | 0.82 | 2.16 | 3.13 |
| | UL | 22.44 | 0.00 | 70.13 | 7.43 | 92.87 | 0.81 | 2.97 | 3.35 |
| | N | 44.87 | 2.04 | 1.94 | 51.14 | 94.77 | 0.82 | 1.44 | 2.97 |
| Secondary | E | 84.89 | 6.80 | 0.79 | 7.52 | 54.72 | 10.58 | 8.04 | 26.66 |
| | US | 31.73 | 31.14 | 8.13 | 28.99 | 53.43 | 10.61 | 8.55 | 27.41 |
| | UL | 13.55 | 0.00 | 63.21 | 23.24 | 51.88 | 10.50 | 9.57 | 28.05 |
| | N | 14.11 | 13.29 | 6.88 | 65.71 | 52.67 | 10.62 | 8.86 | 27.85 |
| Tertiary | E | 71.65 | 1.83 | 0.13 | 26.39 | 8.19 | 0.81 | 1.04 | 89.95 |
| | US | 18.18 | 8.26 | 27.51 | 46.05 | 7.84 | 0.80 | 1.35 | 90.01 |
| | UL | 15.36 | 0.00 | 63.99 | 20.65 | 8.62 | 0.81 | 1.91 | 88.66 |
| | N | 1.78 | 0.65 | 0.12 | 97.45 | 6.59 | 0.79 | 0.96 | 91.66 |

Source: Current Population Survey and authors' calculations.
Notes: Average 1-month and 12-month transition probabilities between hidden states.

the degree of dynamism. It measures the average annual flows per capita between the observed labor market states of employment ($E$), unemployment ($U$), and non-participation ($N$). It is a single summary statistic of the incidence of turnover that is comparable across segments.[20]

The first line of the table shows the relative size of the segments. The majority of the population, 54 percent on average, is part of the primary market. The secondary market is the smallest and consists of 14 percent of the population. The remaining one-third is in the tertiary sector.

### Primary sector

The primary sector is characterized by a low unemployment rate, high labor force participation rate (LFPR), and, consequently, a high employment-to-population (EPOP). Moreover, at 0.5 flows per person per year, turnover in the primary sector is half that in the overall labor market. The blue bars in Figure 4 show the composition of these flows. They help put the low unemployment rate and high EPOP in the primary sector in context.

Labor market frictions are almost irrelevant for primary sector workers who constitute more than half of the population. These workers are almost always employed and they very rarely experience unemployment. When they become unemployed, it is generally for a very short period, because their job-finding rates are much higher than those of others. This can be seen by comparing the 1-month flow rates from $US$ and $UL$ to $E$ in Table 7 for the three sectors.

[20] The estimated flow rates, $q_{l,l',t}$, on which this measure is based are shown in Figures B.7 through B.12.

They also seem to seamlessly move from non-participation to employment and vice versa, unlike workers in the secondary and tertiary sectors for whom these transitions often tend to involve a period of unemployment. In many ways one can think of the labor market experiences of these primary-sector workers as being captured well by those in standard Real Business Cycle (RBC) models (Cooley and Prescott, 1995).

### Secondary sector

The secondary sector is almost the polar opposite of the primary, except for the fact that it also has a high LFPR. The unemployment rate in the secondary sector is more than ten times higher than in the primary sector and almost four times that of the labor market as a whole. Most notably, workers in the sector seem to be in a constant state of flux, as reflected by their flows per capita being six times higher than that of those in the primary sector. As Figure 4 shows, flow rates in the secondary sector are elevated for all six types of flows.

Contrary to workers in the primary sector, labor market frictions are very relevant for workers in the secondary sector. Their labor market experience is characterized by intermittent periods of employment. They frequently move between labor market states and experience unemployment and non-participation spells very often. Because of the importance of labor market frictions for the outcomes of workers in the secondary sector, search models (like Mortensen and Pissarides, 1994) are most applicable to this segment of the labor market.

### Tertiary sector

Only 9 percent of the persons in the tertiary sector participate in the labor market. Those that do have a high unemployment rate of 20 percent. This is mostly caused by those entering the labor force looking for a job. This is different than for the secondary tier, in which job-loss is an important reason for unemployment. Despite similar unemployment rates, secondary and tertiary sectors seem to differ in source of unemployment fluctuations.

### Business cycle fluctuations

So far, we have emphasized the marked differences between the average outcomes of the market segments. In addition to these long-run differences, they also have very distinct business cycle properties. This can be seen from Table 8. It shows the standard deviation of HP-filtered data for each labor market segment. The most striking thing that jumps out from this table is that the secondary market exhibits more volatility for all aggregates we consider. Both the unemployment and labor force participation rates are more than 5 times as volatile in the

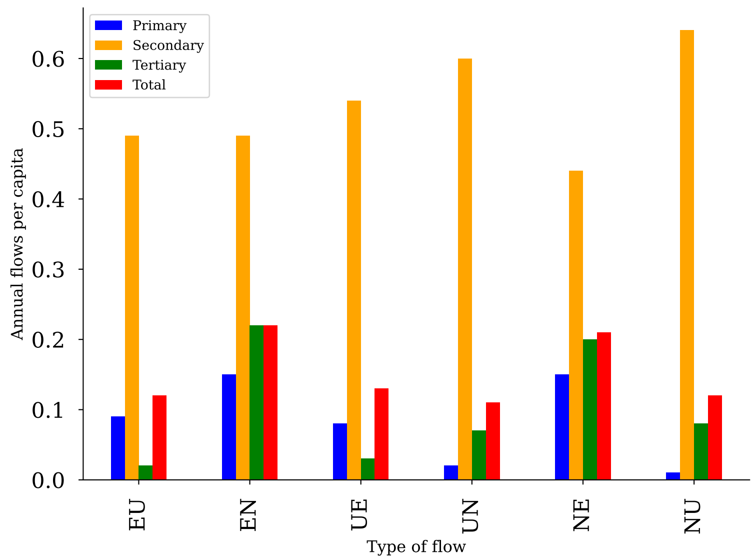



Figure 4: Average annual flows per capita by origin and destination.

*Source:* CPS and authors' calculations.
*Notes:* Averages taken over 1980-2021.

Table 8: Business cycle statistics by labor market segment

| measure | statistic | Primary | Secondary | Tertiary |
|---|---|---|---|---|
| Unemployment rate | $\sigma(x)$ | 0.52 | 2.58 | 2.48 |
| | $\rho(x_t, x_{t-1})$ | 0.71 | 0.78 | 0.81 |
| | $\rho(x_t, Y_t)$ | -0.75 | -0.63 | -0.50 |
| Labor-force participation rate | $\sigma(x)$ | 0.20 | 1.10 | 0.34 |
| | $\rho(x_t, x_{t-1})$ | 0.61 | 0.81 | 0.67 |
| | $\rho(x_t, Y_t)$ | 0.26 | -0.27 | -0.12 |
| Employment-to-population ratio | $\sigma(x)$ | 0.62 | 1.99 | 0.37 |
| | $\rho(x_t, x_{t-1})$ | 0.67 | 0.73 | 0.72 |
| | $\rho(x_t, Y_t)$ | 0.70 | 0.49 | 0.23 |
| Flows per capita | $\sigma(x)$ | 0.06 | 0.13 | 0.02 |
| | $\rho(x_t, x_{t-1})$ | 0.51 | 0.54 | 0.36 |
| | $\rho(x_t, Y_t)$ | -0.61 | -0.20 | -0.11 |

Source: Current Population Survey and authors' calculations.
Business-cycle variables - $\sigma(x)$: standard deviation of HP-filtered cyclical gap from quarterly seasonally adjusted data. $\rho(x_t, x_{t-1})$: first-order autocorrelation of HP-cyclical gap of variable. $\rho(x_t, Y_t)$: correlation of HP-cyclical gap of variable with that of GDP. HP-filter applied with smoothing parameter of 1600.

Table 9: Contribution to aggregates by segment

| | Primary | Secondary | Tertiary | Total |
|---|---|---|---|---|
| Share of population | 54.46 | 13.75 | 31.79 | 100.00 |
| Unemployment rate | 1.66 | 4.09 | 0.88 | 6.62 |
| Labor-force participation rate | 52.91 | 10.04 | 2.81 | 65.77 |
| Employment-to-population ratio | 51.83 | 7.36 | 2.24 | 61.42 |
| Flows per capita | 0.27 | 0.44 | 0.20 | 0.91 |

Source: Current Population Survey and authors' calculations.
Notes: Average percentage-point contribution by market segment to labor market aggregates over sample period. Flows per capita are annual flows between E,U, and N per person.

secondary sector as in the primary sector. The tertiary sector exhibits notable unemployment volatility but the labor force participation rate varies much less than in the secondary segment.

Table 8 also reports the correlation of the cyclical component of each labor market indicator with the cyclical component of GDP. Similar to the aggregate, the unemployment rate is strongly countercyclical in all segments. Despite being very low on average, the primary sector's unemployment rate has the highest negative correlation with GDP growth. Unlike unemployment, the labor force participation rate displays differential cyclical comovement across segments: it is procyclical in the primary sector, countercyclical in the secondary, and almost acyclical in the tertiary. Consequently, the employment-to-population ratio in the primary sector shows the most pronounced positive comovement with output.

## 4.1 Contributions of segments to aggregates

Now that we have established that the U.S. labor market can be viewed as the sum of three very different segments, the next question is how these three distinct parts add up to the whole, i.e. to the labor market aggregates published by the BLS. Here, we answer this question by looking at the contributions of the three market tiers to the averages of and fluctuations in these aggregates.

Table 9 shows these contributions. For example, if we consider the first element of the third row, the primary segment accounts for 53 percentage points of the 66 percent average LFPR in the U.S. over our sample period. The secondary sector, which covers 14 percent of the population, only accounts for about a sixth of the labor supply. The most striking result in the table is that the secondary sector accounts for more than 4 percentage points out of the 6.6 percent average unemployment rate. Thus, one-seventh of the labor force accounts for 60 percent of unemployment. This is the same 14 percent of the population that accounts for 47 percent of gross flows in the economy.

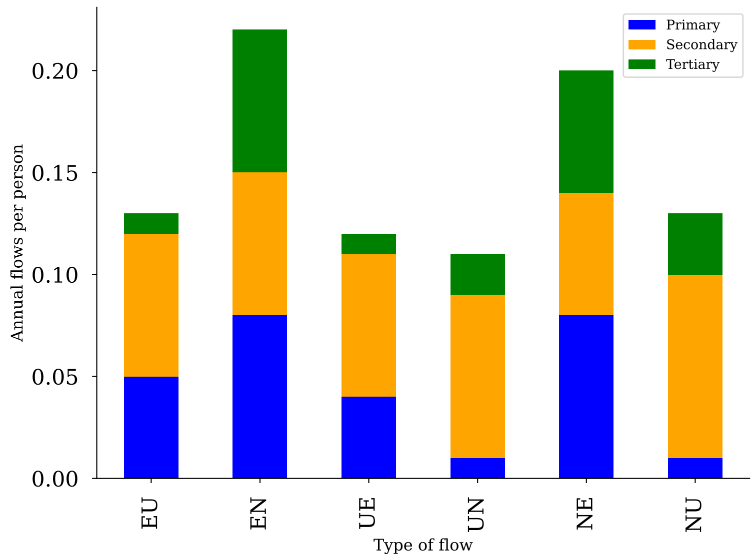



Figure 5: Composition of aggregate flows per person by type of flow and market.

*Source:* CPS and authors' calculations.
*Notes:* Averages taken over 1980-2021.

The importance of the secondary sector for the dynamism of the U.S. labor market is illustrated in Figure 5. It shows that the share of the secondary sector is the highest for flows between unemployment ($U$) and non-participation ($N$). This reflects that the secondary tier is made up of workers at the margin of the labor market. The primary and secondary segments account for almost all job-loss and job-finding in the economy. Interestingly, flows between employment and non-participation are relatively more evenly distributed across different segments.

The secondary sector not only disproportionately contributes to the aggregate unemployment rate, fluctuations in this segment also carry an outsized weight in terms of fluctuations in the labor market aggregates. This can be seen from Table 14. The rows labeled with "$\sigma^2(\Delta x_t)$ " and "$\sigma^2(\Delta_{12} x_t)$" show the contributions of changes in the composition of the population (Share Total) and fluctuations in the labor market aggregates in each segment (Shift) to the monthly and 12-month changes in the aggregates respectively. For example, the second row shows that the secondary market contributes 0.073 to the 0.154 variance of monthly changes in the unemployment rate. We report the statistics for both the monthly and 12-month changes, because the latter exclude seasonal fluctuations.

The shift part in the secondary sector contributes well beyond the 14 percent population weight of the segment to fluctuations in all labor market aggregates considered. The share contributed is higher for monthly changes than for 12-month changes. Thus, the secondary tier absorbs more than its share of seasonal fluctuations and annual fluctuations in the labor market.[21]

## 4.2 Contributions of segments to long-run trends

Over the past four decades the U.S. labor market has seen some profound trends. Our analysis provides a unique perspective on the sources of these trends by examining their drivers in light of the dual labor market structure we uncovered.

The trend decline in the unemployment rate in the US is well known (Shimer, 1999). The origin of this decline is the stark moderation in the incidence of unemployment which declined by more than 50% from 1980s to 2020s as evident by the job-loss and job destruction rates as shown by Davis *et al.* (2010) and Crump *et al.* (2019). We find that this trend is mostly due to the decline in the employment-to-unemployment transition rates in the secondary market. This rate declined from about 10% to 5% in the last 40 years. Moreover, the inflow rate into unemployment from nonparticipation went down from around 20% to 10% in the secondary segment.[22]

The U.S. economy is known to be one of the most fluid labor markets in the world. However, it has been experiencing a notable decline in labor market dynamism as first documented by Davis *et al.* (2007). The decline in dynamism is evident in many different labor market statistics such as job creation, job-to-job transitions and declining business formation. Similar to other trends it is accounted by mostly by the changes in the secondary sector. While there is also some decline in the tertiary sector, it is notable that flows per capita in the primary sector remained largely unchanged at 0.5 in the primary sector over the last 40 years.

Finally, the labor force participation rate has been trending down since its peak in late 1990s. This decline is the result of a shift of the population from the primary to the tertiary sector over time consistent with the aging of the baby boom cohort.

## 4.3 Implications

The large differences in the average outcomes and cyclical sensitivity of the labor market tiers have important implications for the assessment of the costs of business cycles and unemployment as well as for how to evaluate the role and optimal design of unemployment insurance and stabilization policies.

[21] This can also be seen from Figures B.5 and B.6, that plot the time series for the sector-specific labor market aggregates.

[22] See Figures B.9 and B.12 for the relevant time series.

Most studies of the costs of unemployment base their estimates on average incidence rates of unemployment across the population (e.g. Krusell *et al.* , 2010). However, our analysis suggests that these are not the relevant metrics to consider since the unemployment cost of business cycles is disproportionately borne by those in the secondary sector. A proper quantification of this cost should take this inequality into account and distinguish between the costs for workers in different labor market tiers. This insight also reveals that the unemployment insurance system can be thought of as a transfer from those in the primary and tertiary segments to workers in the secondary segment for absorbing a large part of aggregate economic risk over the seasons and business cycle.

The evidence in this section is also relevant for the discussion of policies that aim to stabilize labor-market fluctuations in the short-run. The focus of most of these policies is to maintain unemployment at or around Friedman's (Friedman, 1968) natural rate of unemployment. Because of the different degrees of business-cycle sensitivity across the market segments in our DLM representation, it is important for the implementation of such policies to identify who is in these segments and pay particularly close attention to those in the tier that is most cyclically sensitive. That tier is the secondary market.

Crucial for the two arguments above is the implicit normative assessment that being part of the secondary tier of the labor market is undesirable. This is at the core of many discussions of the DLM Hypothesis. To better understand the context for this value judgment, it is important to analyze the potential sources of the segmentation of the labor market that we quantified. This is what we do in the next section.

# 5 Reasons for labor market segmentation

The undesirability of being part of the secondary labor market segment comes from the implication of DLM that jobs in the primary tier generally pay high wages, come with benefits, offer potential for job advancement, and provide job security. While jobs in the secondary tier have high turnover, pay low wages, come with limited benefits, offer few career opportunities, and provide little job security (Piore, 1970).

In this section we show that the market segments we identified indeed have these properties. We construct our evidence by matching the estimated individual-level posterior probabilities of market-segment membership with data from the CPS on demographic characteristics, industry and occupation of employment, as well as tenure and earnings.[23]

Our ability to link estimated labor market segment membership with all observables in the

[23] One can consider this as a verification of overidentifying restrictions.

CPS also allows us to explore potential reasons for why the dual labor market structure we observe emerged and has persevered. This is important, because the early research on the DLM hypothesis was criticized for not coming to an agreement on these reasons.

Many causes of the segmentation have been emphasized by studies on the DLM. They can broadly be categorized into five themes: (*i*) Life-cycle career choices, (*ii*) discrimination, (*iii*) insider-outsider structure due to labor-market institutions and unionization, (*iv*) efficiency wage theory, and (*v*) labor demand fluctuations.

It might be tempting to interpret these five themes as mutually exclusive and to expect to be able to run a "horse-race empirical analysis" to determine which one shapes the U.S. labor market. However, this would be misguided. These five mechanisms are closely intertwined. For example, the insider-outsider structure of certain parts of the U.S. labor market emerged along racial and gender lines.[24] Saint-Paul (1997) provides a link between efficiency wages, labor adjustment costs, and turnover in the wake of labor-demand fluctuations. Instead, we provide a set of facts that shine a light on the relative importance of each of the five main causes for labor-market segmentation considered in the literature.

**Life-cycle career choices**

Several papers emphasize how individuals learning about their own ability can result in some workers having a sequence of many short employment spells, while others find their comparative advantage and a good match early on in their careers and end up having much longer tenure (Morchio, 2020; Pries, 2004; Pries and Rogerson, 2021). At the core of each of these studies are different permutations of Jovanovic (1979).

Our results show a very distinct life-cycle pattern of labor-market segment membership. This can be seen from Figure 6. It shows the share of the population of a certain age that is part of each market segment for six cohorts of the U.S. population. All six cohorts show the same broad life-cycle patterns: Those younger than 25 are underrepresented in the primary sector and disproportionately work in the secondary sector. In their early twenties there is a gradual transition to the primary sector for most of them.

This suggests that, consistent with theory of life-cycle career choices, a large part of employment in the secondary sector is associated with early-career jobs. Table 10 provides some perspective on the importance of young persons for the secondary segment of the market. Those age 16 to 24 make up about a fifth of our sample and account for one third of those in the

[24]See Hill (1996) for discussion of discrimination by labor unions. Ashenfelter (1972) provides evidence of racial and gender discrimination by labor unions in the '60's and '70's. His evidence indicates that discrimination by unions was less than that in the labor market overall.

secondary tier.

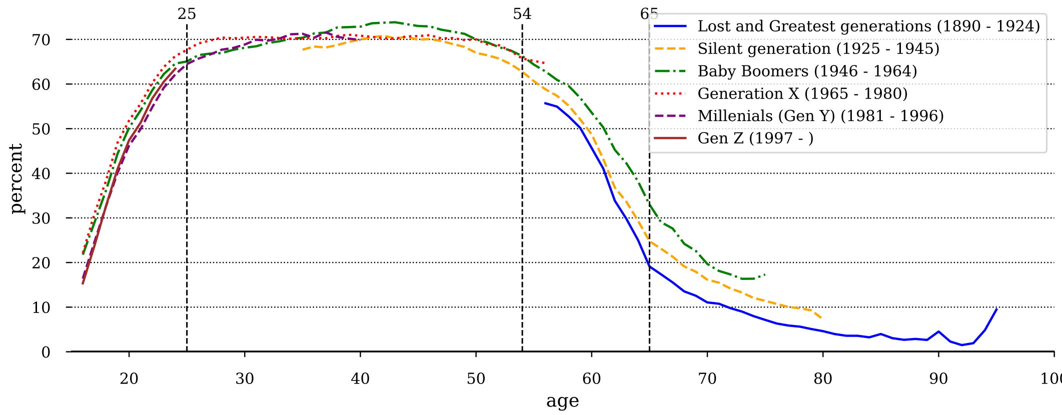
However, panel (b) of the figure also shows that the share of the prime-age population in the secondary tier of the labor market is of the same magnitude as that of the overall population. About one in every eight prime-age persons in the U.S. are part of the secondary segment of the labor market. This share has slowly increased across cohorts over time. Table 10 shows that, even though they are underrepresented in the secondary market, prime-age persons still make up the bulk of it.

Of course, job choice is only one part of early-career decisions. The choice of education is the other part. Young people are likely to have jobs in the secondary segment while they work on their education that provides them access to jobs in the primary sector later on in their career. Table 10 shows that this mechanism is supported in the data. Those with a college education are overrepresented in the primary sector. While those with a high school education make up a disproportionate part of the secondary sector. However, one in every five people in the secondary sector has a college education. So, even though education seems to matter, the bifurcation by education into the primary and secondary tiers is not as stark as one might expect.

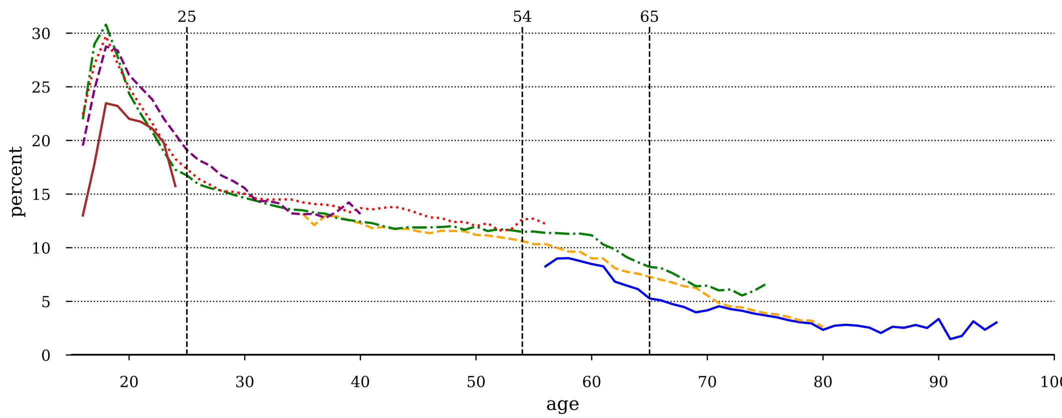
**Discrimination**

Early work on the DLM Hypothesis focused on discrimination. According to this view, the cause of duality is employer's discrimination against women, young adults, racial minorities, and immigrants (Doeringer and Piore, 1970; Berger *et al.* , 1980; Dickens and Lang, 1985). Consistent with this view, we find racial and ethnic minorities and non-naturalized immigrants are overrepresented in the secondary tier of the labor market compared to the primary tier as well as to the overall population as summarized in Table 10. Women make up a larger share of the secondary sector than the primary sector. Most notable for women is that they make up more than 60 percent of the tertiary sector. This largely reflects the low participation rate of women compared to men.

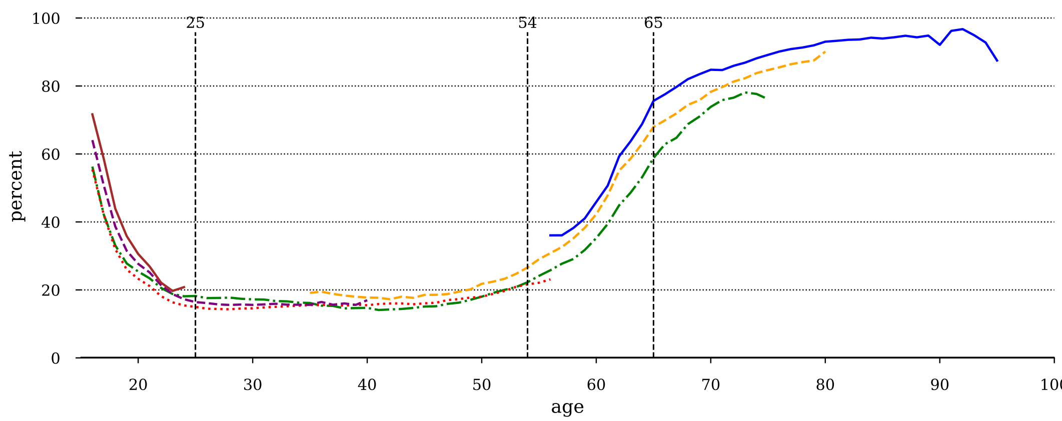
While the evidence in the Table 10 is consistent with discrimination being a source of segmentation, membership of one or more of the aforementioned groups only explains a small fraction of the cross-sectional variation in segment membership. This can be seen from Table 11. It provides the results from regressions of the individual-level posterior probabilities of segment membership on a set of demographic characteristics. The estimated coefficients are in line with the results from Table 10. Women, young adults, and racial and ethnic minorities are all less likely to be in the primary sector than their counterparts. The $R^2$s for the three regressions, however, show that these demographic characteristics explain only a small part of




(a) Primary sector




(b) Secondary sector




(c) Tertiary sector

Figure 6: Segment share by cohort as a function of age

*Source:* CPS and authors' calculations.

Table 10: Composition of market segment by demographic group

| Topic | Segment<br>Group | Primary | Secondary | Tertiary | Total |
|---|---|---|---|---|---|
| Sex | Male | 54.8 | 49.9 | 38.0 | 48.5 |
| | Female | 45.2 | 50.1 | 62.0 | 51.5 |
| Race | White | 82.0 | 74.4 | 81.0 | 80.6 |
| | Black | 11.7 | 18.7 | 12.8 | 13.0 |
| | Other | 6.3 | 6.9 | 6.2 | 6.3 |
| Ethnicity | Not hispanic | 87.6 | 82.8 | 88.8 | 87.4 |
| | Hispanic | 12.4 | 17.2 | 11.2 | 12.6 |
| Age | 16-24 | 16.8 | 33.9 | 21.3 | 20.6 |
| | 25-54 | 68.7 | 53.0 | 29.5 | 53.5 |
| | 55 and over | 14.5 | 13.1 | 49.2 | 25.9 |
| Education | High school or less | 41.9 | 56.7 | 62.9 | 50.5 |
| | Some college | 25.2 | 23.4 | 19.2 | 23.0 |
| | College degree or higher | 33.0 | 19.9 | 18.0 | 26.5 |
| Citizenship | Born in U.S. | 85.2 | 82.7 | 86.0 | 85.1 |
| | Naturalized citizen | 6.3 | 5.4 | 6.2 | 6.1 |
| | Not a citizen | 8.5 | 11.9 | 7.8 | 8.7 |
| Unionization | No union coverage | 85.6 | 88.3 | 87.3 | 86.0 |
| | Covered by union but not a member | 1.7 | 1.4 | 1.5 | 1.6 |
| | Member of labor union | 12.8 | 10.3 | 11.2 | 12.4 |

Source: Current Population Survey and authors' calculations.
Notes: *Some college* denotes some college or associates degree. *Born in U.S.* also includes those born abroad to American parents and those born in outlying U.S. areas.

Table 11: Regression of segment probabilities on demographic characteristics

| | **Primary** | **Secondary** | **Tertiary** |
|---|---|---|---|
| **Female** | -0.1189 | -0.0053 | 0.1241 |
| | (-466.82) | (-31.095) | (524.19) |
| **16-24** | -0.1157 | 0.0618 | 0.0538 |
| | (-270.32) | (217.65) | (135.34) |
| **55 and over** | -0.3628 | -0.0652 | 0.4280 |
| | (-1164.7) | (-315.35) | (1477.6) |
| **Less than high school** | -0.2279 | 0.0545 | 0.1734 |
| | (-533.61) | (192.34) | (436.53) |
| **High school diploma** | -0.1235 | 0.0378 | 0.0857 |
| | (-302.55) | (139.50) | (225.77) |
| **Some college** | -0.0704 | 0.0275 | 0.0429 |
| | (-172.78) | (101.53) | (113.33) |
| **Black** | -0.0700 | 0.0616 | 0.0084 |
| | (-182.25) | (241.52) | (23.591) |
| **Other** | -0.0579 | 0.0175 | 0.0404 |
| | (-109.71) | (49.964) | (82.314) |
| **Hispanic** | -0.0291 | 0.0391 | -0.0100 |
| | (-72.960) | (147.74) | (-26.990) |
| **R-squared** | 0.1891 | 0.0490 | 0.2305 |

Source: Current Population Survey and authors' calculations.
Notes: Number of observations is 10135696. Time fixed effects are included in all regressions. $t$-statistics reported in parentheses. Dummies are normalized to represent a prime-age college-educated white non-hispanic male as the baseline.

the segmentation of the labor market. For all three regressions the $R^2$s are smaller than 0.25. The one for the tertiary sector is highest, mainly because of the life-cycle patterns we discussed above being captured by the age variables. Most notably, the R-squared for the secondary-tier regression is only 0.049. Even though we find evidence consistent with discrimination, it only accounts for a small portion of the segmentation of the labor market.

### Insider-outsider structure of labor market

Early discussions of duality in the U.S. labor market have highlighted different eras of economic distress during which the forces that result in a DLM emerged. For example, Reich *et al.* (1973) focus on the late Nineteenth Century while Berger *et al.* (1980) claim dualism emerged in response to the legislation and labor movements in the wake of the Great Depression during the 1930's. One view is that unionization resulted in insiders and outsiders in the labor market. Insiders getting access to stable high paid careers with benefits, while outsiders do not.[25]

[25] The importance of unions and labor market institutions has been at the heart of the analysis of dualism in European labor markets, e.g. Bentolila *et al.* (2019).

Table 12 shows that those in the primary sector do have more stable and higher paid jobs. The top two measures in the table pertain to job stability. Those in the secondary sector switch from job to job twice as frequently as their counterparts in the primary sector. They also have much shorter tenure than those in the primary sector.[26] Those in the primary sector are mostly full-time employed, i.e. 35 hours a week or higher, while more than half the employed respondents in the secondary and tertiary sectors report working less than 35 hours a week.

This lower number of hours is reflected in usual weekly earnings. The fourth measure in Table 12 reports the distribution of relative usual weekly earnings measured as the percent deviation of an individuals' usual weekly earnings from the median usual weekly earnings in the year. Half of those employed in the secondary segment make 45 percent less than median earnings. This number is similar for those working that are part of the tertiary sector. The difference in weekly earnings is not all due to the fact that those in the secondary and tertiary sectors tend to work less hours. The last measure in Table 12 shows that median hourly earnings in the primary sector are about 30 percent higher than those in the other two tiers.

These differences between the types of jobs in the primary and secondary sectors in terms of stability and earnings are consistent with the DLM hypothesis. They could reflect that the primary sector is made up of insiders while the secondary sector consists of outsiders. However, we find only limited importance of unionization for the dualism of the U.S. labor market. This can be seen from the part of Table 10 related to "unionization". In our sample, 12.8 percent of those in the primary tier report to be members of a union, while only a slightly lower percentage in the secondary tier, i.e. 10.3, does so. Moreover, our results cover 1980-2021 and indicate that the dualism of the U.S. labor market was just as pronounced in the 1980's as in the 2010's. However, union coverage of the U.S. payroll employed halved during that period. This suggests that unions and labor movements more generally, though they might possibly play a role, are not the most important factor driving dualism in the U.S. labor market.

**Efficiency wages**

Insider-outsider effects are only one possible explanation for the differences between jobs in the primary and secondary sectors reported in Table 12. Several studies have emphasized that segmentation of the labor market can be the equilibrium outcome in the presence of market imperfections rather than institutionalized in terms of legislation of unionization. In particular, dualism can emerge when workers' effort on jobs in the primary sector is hard to monitor and

[26] One interesting result is the right tail of the tenure distribution for those in the tertiary sector. This suggests that there is a substantial number of respondents in the CPS that are out the labor force for different reasons and then return to their former employer.

Table 12: Turnover, tenure, hours worked, and earnings

| measure | statistic | Primary | Secondary | Tertiary | Total |
|---|---|---|---|---|---|
| J2J rate | mean | 2.1 | 4.5 | 3.3 | 2.4 |
| Tenure | 10th percentile | 0.5 | 0.2 | 0.2 | 0.4 |
| | 25th percentile | 1.9 | 0.5 | 0.5 | 1.5 |
| | median | 5.0 | 1.8 | 2.0 | 4.0 |
| | 75th percentile | 11.0 | 5.0 | 7.0 | 10.0 |
| | 90th percentile | 20.0 | 12.0 | 20.0 | 20.0 |
| Weekly hours | 10th percentile | 20 | 15 | 10 | 20 |
| | 25th percentile | 37 | 20 | 18 | 32 |
| | median | 40 | 32 | 30 | 40 |
| | 75th percentile | 40 | 40 | 40 | 40 |
| | 90th percentile | 42 | 40 | 40 | 40 |
| Weekly earnings | 10th percentile | -57.6 | -79.6 | -86.2 | -66.0 |
| | 25th percentile | -32.0 | -66.7 | -73.3 | -39.9 |
| | median | 8.3 | -45.5 | -44.0 | 0.0 |
| | 75th percentile | 69.5 | -11.0 | 1.0 | 60.3 |
| | 90th percentile | 153.6 | 47.7 | 68.9 | 142.9 |
| Hourly earnings | 10th percentile | -38.4 | -47.5 | -48.3 | -40.9 |
| | 25th percentile | -22.3 | -37.9 | -38.7 | -26.9 |
| | median | 6.2 | -23.1 | -23.1 | 0.0 |
| | 75th percentile | 52.0 | 4.3 | 7.1 | 44.4 |
| | 90th percentile | 113.3 | 50.4 | 59.8 | 105.8 |

Source: Current Population Survey and authors' calculations.
Notes: *Tenure* Percent of employed that change employers the next month for those who responded to this question (starting in 1994). *Weekly hours* Usual weekly hours worked on all jobs. *Weekly and Hourly earnings* percent deviation from annual median weekly and hourly earnings.

on those in the secondary sector it is not. This results in an efficiency wage structure in the former and a competitive wage structure in the latter (Bulow and Summers, 1986; Albrecht and Vroman, 1992; Saint-Paul, 1997). The equilibrium outcome is higher job stability and wages in the primary sector than in the secondary one.

Our results provide two pieces of support for this efficiency wage theory of dualism. First of all, occupations with higher shares of workers in the primary sector tend to be high-skilled service occupations where effort is hard to monitor and efficiency-wage considerations are likely to play a role in compensation. This can be seen from Panel (a) of Figure 7. The evidence for industries, shown in Panel (b) of the same figure, is not as clear-cut.[27]

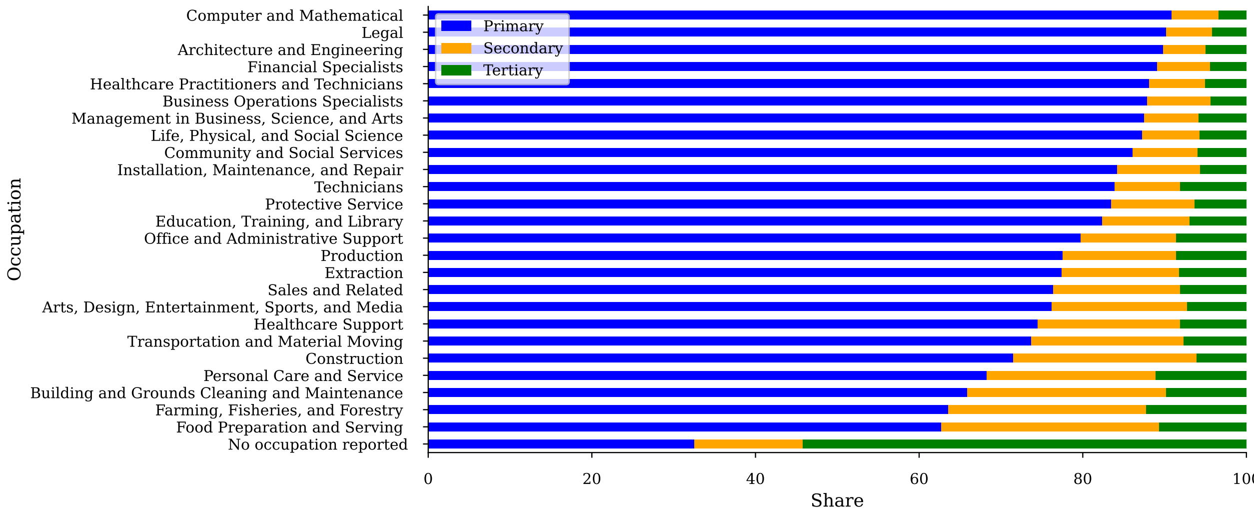
The second piece of evidence, in line with that provided by Dickens and Lang (1985), is that there are significant differences in wage dynamics across the labor market segments. Jobs in the primary sector pay both a higher return to schooling as well as to experience. We show this by runing a generalized Mincer (1974) regression in which we include segment-specific coefficients. These coefficients are those for years of schooling, experience, and experience squared multiplied by the posterior probability that an individual is part of the segment. The results from these regressions for the whole sample, as well as by gender, are reported in Table 13. The return to a year of schooling is about 1.3 percentage points higher in the primary sector than in the secondary and tertiary ones. This is true for both men and women. Returns to experience are also higher in the primary sector, though they decline faster as a person ages.

**Demand fluctuations**

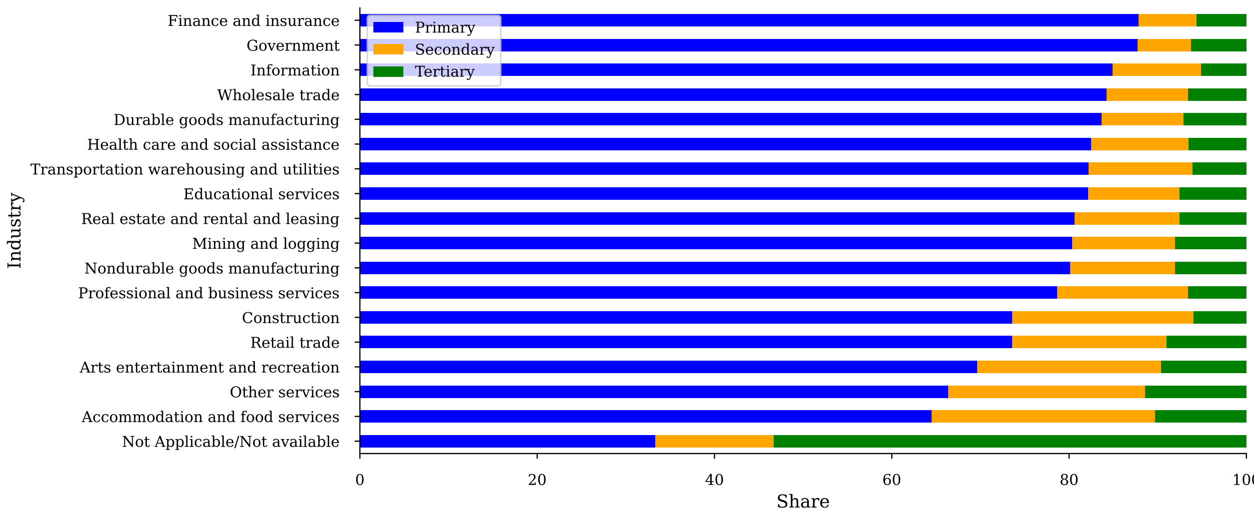
Another potential reason for duality is that it allows workers and firms to organize in a way that insulates jobs that involve match-specific capital from being resolved in response to negative economic shocks. Piore calls this the endogenous "response to flux and uncertainty" (See Berger *et al.* , 1980, Chapter 2). Saint-Paul (1997) illustrates the same intuition in the first figure of his book in the context of labor-adjustment costs due to efficiency wages.

Most discussions of this channel emphasize this mechanism in the context of business cycle fluctuations. However, the non-seasonally-adjusted nature of the individual-level CPS data we use reveals that something similar is true at seasonal frequencies. The secondary segment of the labor market absorbs the bulk of the seasonal fluctuations. Table 14 shows the contributions of the market segments to aggregate labor-market fluctuations. It splits the variance of monthly and 12-month changes in labor market aggregates into parts due to changes in the shares of persons in the market segment and shifts in the aggregates at the segment-level. The 1-month

[27] Efficiency-wage theory is most often analyzed by looking at inter-industry wage differentials (e.g Krueger and Summers, 1988) rather than at the occupational level.




(a) Occupation




(b) Industry

Figure 7: Segment distribution by industry and occupation

*Source:* CPS and authors' calculations.

*Notes:* Industries are based 2-digit NAICS codes and occupations on 2-digit 2010 SOC codes.

Table 13: Mincer regressions separated by market segment

| | **Total** | **Male** | **Female** |
|---|---|---|---|
| **Dep. Variable** | hourly wage (log) | hourly wage (log) | hourly wage (log) |
| **Years of schooling - Primary** | 0.0706 | 0.0655 | 0.0855 |
| | (655.10) | (449.61) | (557.05) |
| **Years of schooling - Secondary** | 0.0568 | 0.0514 | 0.0715 |
| | (369.57) | (240.83) | (338.45) |
| **Years of schooling - Tertiary** | 0.0598 | 0.0528 | 0.0764 |
| | (265.64) | (158.77) | (260.88) |
| **Experience - Primary** | 0.0344 | 0.0397 | 0.0278 |
| | (508.85) | (415.55) | (306.68) |
| **Experience - Secondary** | 0.0214 | 0.0278 | 0.0177 |
| | (126.52) | (114.30) | (79.853) |
| **Experience - Tertiary** | 0.0202 | 0.0299 | 0.0203 |
| | (65.468) | (57.121) | (54.434) |
| **Experience$^2$ - Primary** | -0.0006 | -0.0006 | -0.0005 |
| | (-396.17) | (-318.45) | (-235.07) |
| **Experience$^2$ - Secondary** | -0.0003 | -0.0004 | -0.0002 |
| | (-78.785) | (-72.415) | (-46.353) |
| **Experience$^2$ - Tertiary** | -0.0003 | -0.0004 | -0.0003 |
| | (-40.580) | (-39.118) | (-36.477) |
| **No. Observations** | 2986456 | 1452501 | 1533955 |
| **R-squared** | 0.2632 | 0.2926 | 0.2748 |
| **Effects** | Time | Time | Time |

Source: Current Population Survey and authors' calculations.
Notes: Year fixed effects are included in all regressions. $t$-statistics reported in parentheses. Experience is defined as age minus years of schooling minus six.

Table 14: Shift-share analysis of changes in labor market aggregates

| | | Sum Total | Share Total | Shift Primary | Secondary | Tertiary |
|---|---|---|---|---|---|---|
| Unemployment rate | $\bar{\Delta x_t}$ | -0.0929 | 0.0023 | -0.0240 | -0.0577 | -0.0135 |
| | $\sigma^2(\Delta x_t)$ | 0.1546 | 0.0123 | 0.0516 | 0.0746 | 0.0161 |
| | $\sigma^2(\Delta_{12} x_t)$ | 1.1299 | 0.1734 | 0.3628 | 0.4652 | 0.1285 |
| Labor-force participation rate | $\bar{\Delta x_t}$ | 0.0001 | 0.0373 | -0.0061 | -0.0219 | -0.0092 |
| | $\sigma^2(\Delta x_t)$ | 0.1602 | 0.0329 | 0.0133 | 0.0574 | 0.0565 |
| | $\sigma^2(\Delta_{12} x_t)$ | 0.1636 | 0.1102 | 0.0093 | 0.0207 | 0.0234 |
| Employment-to-population ratio | $\bar{\Delta x_t}$ | 0.0594 | 0.0335 | 0.0093 | 0.0219 | -0.0053 |
| | $\sigma^2(\Delta x_t)$ | 0.1802 | 0.0340 | 0.0380 | 0.0689 | 0.0393 |
| | $\sigma^2(\Delta_{12} x_t)$ | 0.6837 | 0.2736 | 0.1655 | 0.1773 | 0.0672 |
| Flows per capita | $\bar{\Delta x_t}$ | -0.0060 | -0.0004 | -0.0005 | -0.0037 | -0.0014 |
| | $\sigma^2(\Delta x_t)$ | 0.0081 | 0.0001 | 0.0025 | 0.0032 | 0.0022 |
| | $\sigma^2(\Delta_{12} x_t)$ | 0.0022 | 0.0002 | 0.0008 | 0.0007 | 0.0005 |

Source: Current Population Survey and authors' calculations.
Notes: Contributions to average changes and the variance of 1-month and 12-month changes.

changes include seasonality while the 12-month changes do not. Even though the secondary sector only makes up 14 percent of the population, it accounts for more than 40 percent of the 1- and 12-month fluctuations in the unemployment rate as well as for more than 30 percent of the fluctuations in the EPOP.

This points to dualism in the U.S. labor market persisting to organize the division of labor in the face of labor market imperfections in a way that minimizes adjustment costs in response to predictable seasonal as well as unpredictable business cycle fluctuations. Our results likely underestimate the importance of this channel because the CPS data we use is collected at the monthly frequency. It does not include the high-frequency turnover in the labor market.[28]

What remains an open question is what determines why particular workers end up in different labor market segments. The model in Albrecht and Vroman (1992) emphasizes heterogeneity in the value of non-employment as a major determinant of this. To make progress on this question and test this hypothesis, better measures of this value at the individual level would be required.

# 6 Conclusion

The dynamics of the stocks and flows in the U.S. labor market are well captured by a DLM with a tertiary sector made up of those who participate infrequently. This interpretation provides

[28] Like, for example, the Starbucks barista who only works peak-demand shifts during workdays from 7.30am through 10.30am and spends the rest of the day studying for her law degree.

a parsimonious framework within which many aspects that have puzzled labor- and macroeconomists can be interpreted. The three market segments can be disentangled using an unsupervised machine learning method that involves the estimation of an HMM with identifying inequality constraints on the transition probabilities. These restrictions are what ensures that the hidden states we uncover can be interpreted as making up the primary, secondary, and tertiary labor-market tiers. What emerges is a tale of three totally different sub-markets.

Labor market frictions are basically irrelevant for primary sector workers who make up around 55 percent of the population. These workers are almost always employed and they very rarely experience unemployment. They also seamlessly move from non-participation to employment unlike workers in the secondary and tertiary sectors. The secondary sector, which constitutes 14 percent of the population, exhibits high turnover and high unemployment and absorbs most of the short-run fluctuations in the labor market, at both seasonal and business cycle frequencies. Workers in this sector are six times more likely to move between labor market states than those in the primary tier and are 10 more likely to be unemployed than their primary counterparts. The tertiary sector mostly includes workers who are only loosely attached to the labor market and has a very low employment-to-population ratio. These workers mostly experience unemployment when they enter the labor force from nonparticipation but do not share the high job-loss rate of secondary workers.

Because the total labor market is the sum of these three very different parts, average outcomes, which are often used for to quantitatively discipline macroeconomic models of the labor market, are not reflective of the labor market experiences of anyone in the population. Better quantitative analyses should take into account the Macro Heterogeneity in labor market outcomes we uncovered in this paper.

# References

Abowd, John M., and Zellner, Arnold. 1985. Estimating Gross Labor-Force Flows. *Journal of Business and Economic Statistics*, **3**(3), 254–283.

Ahn, Hie Joo, and Hamilton, James D. 2020a. Heterogeneity and Unemployment Dynamics. *Journal of Business and Economic Statistics*, **38**(3), 554–569.

Ahn, Hie Joo, and Hamilton, James D. 2020b (June). *Measuring Labor-Force Participation and the Incidence and Duration of Unemployment*. Working Paper 27394. National Bureau of Economic Research.

Albrecht, James W., and Vroman, Susan B. 1992. Dual Labor Markets, Efficiency Wages, and Search. *Journal of Labor Economics*, **10**(4), 438–461.

Allman, Elizabeth S., Matias, Catherine, and Rhodes, John A. 2009. Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.*, **37**(6A), 3099–3132.

Amisano, Gianni, and Tristani, Oreste. 2019 (Apr.). *Uncertainty Shocks, Monetary Policy and Long-Term Interest Rates*. Finance and Economics Discussion Series 2019-024. Board of Governors of the Federal Reserve System (U.S.).

Andersen, Martin, Dahl, Joachim, Liu, Zhang, and Vandenberghe, Lieven. 2011. Interior-point methods for large-scale cone programming. *Optimization for machine learning*, **5583**.

Ashenfelter, Orley. 1972. Racial Discrimination and Trade Unionism. *Journal of Political Economy*, **80**(3), 435–464.

Barnichon, Regis, and Nekarda, Christopher J. 2012. The Ins and Outs of Forecasting Unemployment: Using Labor Force Flows to Forecast the Labor Market. *Brookings Papers on Economic Activity*, **43**(2 (Fall)), 83–131.

Baum, Leonard E, Petrie, Ted, Soules, George, and Weiss, Norman. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The annals of mathematical statistics*, **41**(1), 164–171.

Baumeister, Christiane, and Hamilton, James D. 2015. Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information. *Econometrica*, **83**(5), 1963–1999.

Bentolila, Samuel, Dolado, Juan J., and Jimeno, Juan F. 2019 (Jan.). *Dual Labour Markets Revisited*. IZA Discussion Papers 12126. Institute of Labor Economics (IZA).

Berger, S., Berger, P.P.S.S., Piore, M.J., Suzanne, B., and Press, Cambridge University. 1980. *Dualism and Discontinuity in Industrial Societies*. Cambridge University Press.

Blanchard, Olivier Jean, Diamond, Peter, Hall, Robert E., and Murphy, Kevin. 1990. The Cyclical Behavior of the Gross Flows of U.S. Workers. *Brookings Papers on Economic Activity*, **1990**(2), 85–155.

Boeschoten, Laura, Filipponi, Danila, and Varriale, Roberta. 2020. Combining Multiple Imputation and Hidden Markov Modeling to Obtain Consistent Estimates of Employment Status. *Journal of Survey Statistics and Methodology*, 01. smz052.

Bulow, Jeremy I, and Summers, Lawrence. 1986. A Theory of Dual Labor Markets with Application to Industrial Policy,Discrimination, and Keynesian Unemployment. *Journal of Labor Economics*, **4**(3), 376–414.

Cain, Glen G. 1975. The Challenge of Dual and Radical Theories of the Labor Market to Orthodox Theory. *The American Economic Review*, **65**(2), 16–22.

Christiano, Lawrence J, Eichenbaum, Martin, and Vigfusson, Robert. 2006 (July). *Assessing Structural VARs*. Working Paper 12353. National Bureau of Economic Research.

Cooley, Thomas F., and Prescott, Edward C. 1995. Economic Growth and Business Cycles. *Chap. 1, pages 1–38 of:* Cooley, Thomas F. (ed), *Frontiers of Business Cycle Research*. Princeton, New Jersey: Princeton University Press.

Costain, James, Jimeno, Juan F., and Thomas, Carlos. 2010 (Apr.). *Employment fluctuations in a dual labor market*. Working Papers 1013. Banco de España.

Crump, Richard K., Eusepi, Stefano, Giannoni, Marc, and Şahin, Ayşegül. 2019 (June). *A Unified Approach to Measuring u*. NBER Working Papers 25930. National Bureau of Economic Research, Inc.

Davis, Steven J., Haltiwanger, John, Jarmin, Ron, and Miranda, Javier. 2007. Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms. *Pages 107–180 of: NBER Macroeconomics Annual 2006, Volume 21*. NBER Chapters. National Bureau of Economic Research, Inc.

Davis, Steven J., Faberman, R. Jason, Haltiwanger, John, Jarmin, Ron, and Miranda, Javier. 2010. Business Volatility, Job Destruction, and Unemployment. *American Economic Journal: Macroeconomics*, **2**(2), 259–287.

Dempster, Arthur P, Laird, Nan M, and Rubin, Donald B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**(1), 1–22.

Dickens, William T., and Lang, Kevin. 1985. A Test of Dual Labor Market Theory. *The American Economic Review*, **75**(4), 792–805.

Doeringer, P.B., and Piore, M.J. 1970. *Internal Labor Markets and Manpower Analysis*. Manpower Administration (Department of Labor).

Elsby, Michael W. L., Hobijn, Bart, Şahin, Ayşegül, and Valletta, Robert G. 2011. The Labor Market in the Great Recession — An Update to September 2011. *Brookings Papers on Economic Activity*, **42**(2 (Fall)), 353–384.

Elsby, Michael W.L., Hobijn, Bart, and Şahin, Ayşegül. 2015. On the importance of the participation margin for labor market fluctuations. *Journal of Monetary Economics*, **72**(C), 64–82.

Feng, Shuaizhang, and Hu, Yingyao. 2013. Misclassification Errors and the Underestimation of the US Unemployment Rate. *American Economic Review*, **103**(2), 1054–70.

Ferraro, Domenico. 2018. The Asymmetric Cyclical Behavior of the U.S. Labor Market. *Review of Economic Dynamics*, **30**(October), 145–162.

Flood, Sarah, King, Miriam, Rodgers, Renae, Ruggles, Steven, and Warren, J. Robert. 2020. *Integrated Public Use Microdata Series, Current Population Survey, Version 8.0 [dataset]*. University of Minnesota.

Friedman, Milton. 1968. The Role of Monetary Policy. *The American Economic Review*, **58**(1), 1–17.

Granger, C. W. J., and Morris, M. J. 1976. Time Series Modelling and Interpretation. *Journal of the Royal Statistical Society. Series A (General)*, **139**(2), 246–257.

Gregory, Victoria, Menzio, Guido, and Wiczer, David G. 2021 (Apr.). *The Alpha Beta Gamma of the Labor Market*. NBER Working Papers 28663. National Bureau of Economic Research, Inc.

Guner, Nezih, Ventura, Gustavo, and Xu, Yi. 2008. Macroeconomic implications of size-dependent policies. *Review of Economic Dynamics*, **11**(4), 721–744.

Hall, Robert E. 1982. The Importance of Lifetime Jobs in the U.S. Economy. *The American Economic Review*, **72**(4), 716–724.

Hall, Robert E., and Kudlyak, Marianna. 2019 (Feb.). *Job-Finding and Job-Losing: A Comprehensive Model of Heterogeneous Individual Labor-Market Dynamics*. NBER Working Papers 25625. National Bureau of Economic Research, Inc.

Hall, Robert E., and Schulhofer-Wohl, Sam. 2018. Measuring Job-Finding Rates and Matching Efficiency with Heterogeneous Job-Seekers. *American Economic Journal: Macroeconomics*, **10**(1), 1–32.

Harrison, Bennett, and Sum, Andrew. 1979. The Theory of "Dual" or Segmented Labor Markets. *Journal of Economic Issues*, **13**(3), 687–706.

Hill, Herbert. 1996. The Problem of Race in American Labor History. *Reviews in American History*, **24**(2), 189–208.

Hornstein, Andreas. 2012 (11). *Accounting for unemployment: the long and short of it*. Working Paper 12-07. Federal Reserve Bank of Richmond.

Hyatt, Henry, and Spletzer, James. 2016. The shifting job tenure distribution. *Labour Economics*, **41**(C), 363–377.

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

Kroft, Kory, Lange, Fabian, Notowidigdo, Matthew J., and Katz, Lawrence F. 2016. Long-Term Unemployment and the Great Recession: The Role of Composition, Duration Dependence, and Nonparticipation. *Journal of Labor Economics*, **34**(S1), S7–S54.

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

Krusell, Per, Mukoyama, Toshihiko, and Şahin, Ayşegül. 2010. Labour-Market Matching with Precautionary Savings and Aggregate Fluctuations. *The Review of Economic Studies*, **77**(4), 1477–1507.

Krusell, Per, Mukoyama, Toshihiko, Rogerson, Richard, and Şahin, Ayşegül. 2017. Gross Worker Flows over the Business Cycle. *American Economic Review*, **107**(11), 3447–3476.

Kudlyak, Marianna, and Lange, Fabian. 2017 (Sept.). *Measuring Heterogeneity in Job Finding Rates among the Non-Employed Using Labor Force Status Histories*. Working Paper Series 2017-20. Federal Reserve Bank of San Francisco.

Levinson, S. E., Rabiner, L. R., and Sondhi, M. M. 1983. An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *The Bell System Technical Journal*, **62**(4), 1035–1074.

Marston, Stephen T. 1976. Employment Instability and High Unemployment Rates. *Brookings Papers on Economic Activity*, **1976**(1), 169–210.

Maruotti, Antonello. 2011. Mixed Hidden Markov Models for Longitudinal Data: An Overview. *International Statistical Review*, **79**(3), 427–454.

McNabb, Robert, and Psacharopoulos, George. 1981. Further Evidence of the Relevance of the Dual Labor Market Hypothesis for the U. K. *The Journal of Human Resources*, **16**(3), 442–448.

Menzio, Guido, and Shi, Shouyong. 2011. Efficient Search on the Job and the Business Cycle. *Journal of Political Economy*, **119**(3), 468–510.

Mincer, Jacob A. 1974. *The Human Capital Earnings Function*. NBER. Pages 83–96.

Morchio, Iacopo. 2020. Work Histories and Lifetime Unemployment. *International Economic Review*, **61**(1), 321–350.

Mortensen, Dale T., and Pissarides, Christopher A. 1994. Job Creation and Job Destruction in the Theory of Unemployment. *The Review of Economic Studies*, **61**(3), 397–415.

Otterpohl, Jens R. 2002. Baum-Welch Learning in Discrete Hidden Markov Models with Linear Factorial Constraints. *Pages 1180–1185 of:* Dorronsoro, José R. (ed), *Artificial Neural Networks — ICANN 2002*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Piore, Michael J. 1970. The Dual Labor Market: Theory and Implications. *Pages 55–59 of:* Beer, Samuel H., and Barringer, Richard E. (eds), *The State and the Poor*.

Pries, Michael J. 2004. Persistence of Employment Fluctuations: A Model of Recurring Job Loss. *Review of Economic Studies*, **71**(1), 193–215.

Pries, Michael J, and Rogerson, Richard. 2021. Declining Worker Turnover: The Role of Short Duration Employment Spells. *American Economic Journal: Macroeconomics*.

Reich, Michael, Gordon, David M., and Edwards, Richard C. 1973. A Theory of Labor Market Segmentation. *The American Economic Review*, **63**(2), 359–365.

Saint-Paul, Gilles. 1997. *Dual Labor Markets: A Macroeconomic Perspective*. MIT Press Books, vol. 1, no. 0262193760. The MIT Press.

Schorfheide, Frank. 2005. Learning and Monetary Policy Shifts. *Review of Economic Dynamics*, **8**(2), 392–419.

Sedlacek, Petr. 2016. The aggregate matching function and job search from employment and out of the labor force. *Review of Economic Dynamics*, **21**(July), 16–28.

Shibata, Ippei. 2019 (Dec.). *Labor Market Dynamics: A Hidden Markov Approach*. IMF Working Papers 2019/282. International Monetary Fund.

Shimer, Robert. 1999. Why is the U.S. Unemployment Rate So Much Lower? *Pages 11–74 of: NBER Macroeconomics Annual 1998, volume 13*. NBER Chapters. National Bureau of Economic Research, Inc.

Shimer, Robert. 2005. The Cyclical Behavior of Equilibrium Unemployment and Vacancies. *American Economic Review*, **95**(1), 25–49.

Stock, James H., and Watson, Mark W. 2001. Vector Autoregressions. *Journal of Economic Perspectives*, **15**(4), 101–115.

van den Berg, Gerard J., and van Ours, Jan C. 1996. Unemployment Dynamics and Duration Dependence. *Journal of Labor Economics*, **14**(1), 100–125.

Veracierto, Marcelo. 2008. On the cyclical behavior of employment, unemployment and labor force participation. *Journal of Monetary Economics*, **55**(6), 1143–1157.

Wachter, Michael L. 1974. Primary and Secondary Labor Markets: A Critique of the Dual Approach. *Brookings Papers on Economic Activity*, **5**(3), 637–694.

Welch, Lloyd R. 2003. Hidden Markov Models and the Baum-Welch Algorithm. *IEEE Information Theory Society Newsletter*, **53**(4), 1,10–23.

# A Mathematical and computational details

### E-step: Conditional expectation of the complete-data log-likelihood

The conditional expectation of the complete-data log likelihood function can be derived by considering the expectations of $u_{i,t,l}$ and $v_{i,t,l,l'}$. For the first one, we obtain

$$\begin{aligned}\hat{u}_{i,t,l} &= E\left[u_{i,t,l} \mid \boldsymbol{x}_i; \boldsymbol{\theta}\right] = P\left(\ell_{i,t} = l \mid \boldsymbol{x}_i; \boldsymbol{\theta}\right) \\ &= \frac{P\left(\ell_{i,t} = l \cap \boldsymbol{x}_i; \boldsymbol{\theta}\right)}{P\left(\boldsymbol{x}_i; \boldsymbol{\theta}\right)} = \frac{P\left(\ell_{i,t} = l \cap \boldsymbol{x}_i; \boldsymbol{\theta}\right)}{\sum_{l' \in L} P\left(\ell_{i,t} = l' \cap \boldsymbol{x}_i; \boldsymbol{\theta}\right)}.\end{aligned} \tag{18}$$

Similarly

$$\begin{aligned}\hat{v}_{i,t,l,l'} &= E\left[v_{i,t,l,l'} \mid \boldsymbol{x}_i; \boldsymbol{\theta}\right] = P\left(\ell_{i,t-1} = l \cap \ell_{i,t} = l' \mid \boldsymbol{x}_i; \boldsymbol{\theta}\right) \\ &= \frac{P\left(\ell_{i,t-1} = l \cap \ell_{i,t} = l' \cap \boldsymbol{x}_i; \boldsymbol{\theta}\right)}{P\left(\boldsymbol{x}_i; \boldsymbol{\theta}\right)} \\ &= \frac{P\left(\ell_{i,t-1} = l \cap \ell_{i,t} = l' \cap \boldsymbol{x}_i; \boldsymbol{\theta}\right)}{\sum_{h' \in L} \sum_{h \in L} P\left(\ell_{i,t-1} = h \cap \ell_{i,t} = h' \cap \boldsymbol{x}_i; \boldsymbol{\theta}\right)}.\end{aligned} \tag{19}$$

Here, we can express

$$\begin{aligned}P\left(\ell_{i,t_i+k} = l \cap \boldsymbol{x}_i; \boldsymbol{\theta}\right) &= P\left(x_{i,t_i}, .., x_{i,t_i+k} \cap \ell_{i,t_i+k} = l\right) \\ &\quad \times P\left(x_{i,t_i+k+1}, .., x_{i,t_i+15} \mid \ell_{i,t_i+k} = l\right) \\ &= \alpha_{i,k}\left(l\right) \beta_{i,k}\left(l\right),\end{aligned} \tag{20}$$

where $\alpha_{i,k}\left(l\right)$ is as defined in the main text and

$$\beta_{i,k}\left(l\right) = P\left(x_{i,t_i+k+1}, .., x_{i,t_i+15} \mid \ell_{i,t_i+k} = l\right). \tag{21}$$

Moreover

$$\begin{aligned}P\left(\ell_{i,t_i+k-1} = l \cap \ell_{i,t_i+k} = l' \cap \boldsymbol{x}_i; \boldsymbol{\theta}\right) &= P\left(x_{i,t_i}, .., x_{i,t_i+k-1} \cap \ell_{i,t_i+k-1} = l\right) \\ &\quad \times q_{t_i+k,l,l'} \omega_{x_{i,t_i+k},l',t_+k} \\ &\quad \times P\left(x_{i,t_i+k+1}, .., x_{i,t_i+15} \mid \ell_{i,t_i+k} = l'\right) \\ &= \alpha_{i,k-1}\left(l\right) q_{t_i+k,l,l'} \omega_{x_{i,t_i+k},l',t_+k} \beta_{i,k}\left(l'\right)\end{aligned} \tag{22}$$

This yields that

$$\hat{u}_{i,t_i+k,l} = \frac{\alpha_{i,k}(l)\,\beta_{i,k}(l)}{\sum_{l'\in L}\alpha_{i,k}(l')\,\beta_{i,k}(l')} \tag{23}$$

and

$$\hat{v}_{i,t_i+k,l,l'} = \frac{\alpha_{i,k-1}(l)\,q_{t_i+k,l,l'}\omega_{x_{i,t_i+k},l',t+k}\beta_{i,k}(l')}{\sum_{h'\in L}\sum_{h\in L}\alpha_{i,k-1}(l)\,q_{t_i+k,h,h'}\omega_{x_{i,t_i+k},h',t+k}\beta_{i,k}(l')}. \tag{24}$$

Just like $\alpha_{i,k}(l)$, $\beta_{i,k}(l)$ can be calculated using a recursion.

$$\begin{aligned}
\beta_{i,15}(l) &= 1\text{, and} && (25)\\
\beta_{i,k}(l) &= P\left(x_{i,t_i+k+1},..,x_{i,t_i+15} \mid \ell_{i,t_i+k}=l\right) && (26)\\
&= \sum_{l'} q_{t_i+k+1,l,l'}\beta_{i,k+1}(l') && (27)\\
&\quad \times\left[(1-\eta_{i,t_i+k+1})+\eta_{i,t_i+k+1}\omega_{x_{i,t_i+k+1},l',t_i+k+1}\right]\text{, for } k=0,\ldots,14 && (28)
\end{aligned}$$

is the backward recursion that is part of the Forward-Backward method (BW).

**Property of posterior probabilities**

Let $\{\mathcal{P},\mathcal{S},\mathcal{T}\}$ be the sets of hidden labor market states that are part of the primary and secondary tiers respectively. If there is no mobility between these tiers then it must be the case that for $\mathcal{M}\in\{\mathcal{P},\mathcal{S},\mathcal{T}\}$:

$$\begin{aligned}
P\left(\boldsymbol{x}_i\cap\ell_{i,t}\in\mathcal{M}\right) &= \sum_{l\in\mathcal{T}} P\left(\boldsymbol{x}_i\cap\ell_{i,t}=l\right) && (29)\\
&= \sum_{l\in\mathcal{M}}\sum_{l'\in\mathcal{M}} P\left(\ell_{i,t}=l\mid\ell_{i,t-1}=l'\right)P\left(\boldsymbol{x}_i\cap\ell_{i,t-1}=l'\right) && (30)\\
&= \sum_{l'\in\mathcal{M}}\sum_{l\in\mathcal{M}} P\left(\ell_{i,t}=l\mid\ell_{i,t-1}=l'\right)P\left(\boldsymbol{x}_i\cap\ell_{i,t-1}=l'\right) && (31)\\
&= \sum_{l'\in\mathcal{M}} P\left(\boldsymbol{x}_i\cap\ell_{i,t-1}=l'\right) && (32)\\
&= P\left(\boldsymbol{x}_i\cap\ell_{i,t-1}\in\mathcal{M}\right) && (33)
\end{aligned}$$

Thus, the posterior probability that a person is in a particular segment of the labor market is constant over time when there is no mobility across the labor market tiers.

**M-step: Updated parameter estimates**

In the M-step, the parameters, $\delta_{l,t_i}$, $q_{t_i+k,l,l'}$, and $\omega_{x_{i,t_i+k},l,t_i+k}$, are chosen to maximize

$$\begin{aligned} \ln \mathcal{L} &= \sum_{i=1}^{n} w_i \left\{ \sum_{l \in L} \hat{u}_{i,t_i,l} \ln \delta_{l,t_i} + \sum_{k=1}^{15} \sum_{l' \in L} \sum_{l \in L} \hat{v}_{i,t_i+k,l,l'} \ln q_{t_i+k,l,l'} \right. \\ &\left. + \sum_{k=0}^{15} \eta_{i,t_i+k} \sum_{l \in L} \hat{u}_{i,t_i+k,l} \ln \omega_{x_{i,t_i+k},l,t_i+k} \right\}. \end{aligned} \tag{34}$$

subject to the adding-up constraints

$$\sum_{l} \delta_{l,t} = 1, \text{ for } t = 1, \ldots, T \tag{35}$$

$$\sum_{l}^{\prime} q_{t,l,l'} = 1, \text{ for } t = 1, \ldots, T \text{ and } l \in L, \text{ and} \tag{36}$$

$$\sum_{x \in X} \omega_{x_{i,t_i+k},l,t_i+k} = 1, \text{ for } t = 1, \ldots, T \text{ and } l \in L \tag{37}$$

as well as the additional (in-)equality restrictions we described in Subsection 3.2.

Without the additional identifying (in-)equality constraints, the above maximization problem has a closed-form solution derived in Baum *et al.* (1970); Welch (2003). The implementation of the BW with parameter constraints has been studied extensively (most notably Levinson *et al.* , 1983; Otterpohl, 2002). Under some types of constraints the M-step yields closed-form solutions. But that is not the case for our application. Instead, we rely on numerical methods to maximize the expected complete-data likelihood.

We exploit that the identifying restrictions we impose have two important properties. The first is that they are all contemporaneous in that they impose restrictions on parameters at the same point in time. The second is that they are separated between transition probabilities, $q_{t,l,l'}$, and emission probabilities, $\omega_{x,l,t}$.

This property simplifies the M-step to $3T$ convex maximization problems. To see how this works, define the set $N(t)$ as the individuals $i$ who are respondents in period $t$. Then we can write

$$\ln \mathcal{L} = \sum_{T}^{t=1} \sum_{i \in N(t)} w_i \left\{ \sum_{l \in L} \hat{u}_{i,t,l} \ln \delta_{l,t} + \sum_{l \in L} \sum_{l' \in L} \hat{v}_{i,t,l,l'} \ln q_{t,l,l'} + \sum_{l \in L} \hat{u}_{i,t,l} \ln \omega_{x_{i,},l,t} \right\}.$$

Then, for each month $t$ the M-step involves three maximization problems. The first is to maximize

$$\sum_{l \in L} \hat{u}_{i,t,l} \ln \delta_{l,t}, \tag{38}$$

with respect to the unconditional probabilities (stocks), $\{\delta_{l,t}\}_{l \in L}$, subject to the adding-up constraint (35). This is a well-defined convex problem that solves for the Weighted Analytic Center that can be solved using the algorithm from Andersen *et al.* (2011).

The other two problems also involve solving for the Weighted Analytic Center but subject to more constraints. The transition probabilities in month $t$, $\{q_{l,l',t}\}_{(l,l') \in L \times L}$, in the M-step maximize

$$\sum_{l \in L} \sum_{l' \in L} \hat{v}_{i,t,l,l'} \ln q_{t,l,l'}, \tag{39}$$

subject to (36) and the identifying inequality constraints introduced in Subsection 3.2. This, again can be solved using the algorithm from Andersen *et al.* (2011). The same is true for the emission probabilities, $\{\omega_{x,l,t}\}_{(x,l) \in X \times L}$, which maximize

$$\sum_{l \in L} \hat{u}_{i,t,l} \ln \omega_{x_{i,},l,t}, \tag{40}$$

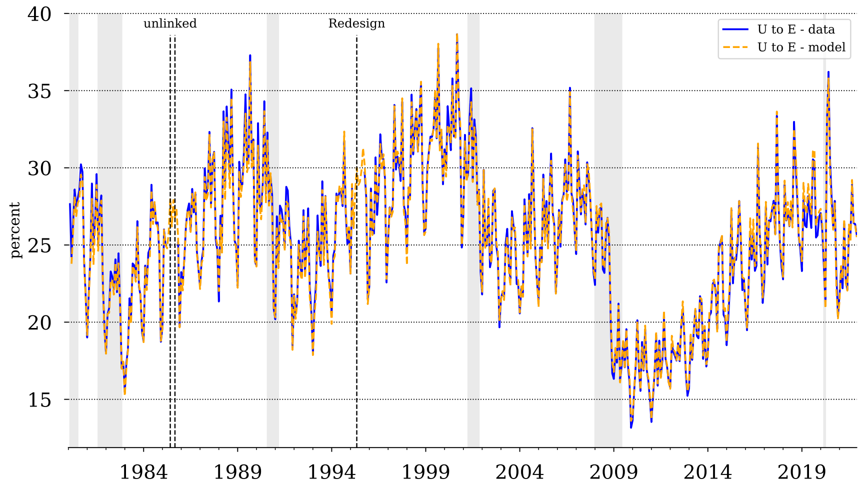
subject to (37).

# B Additional empirical results

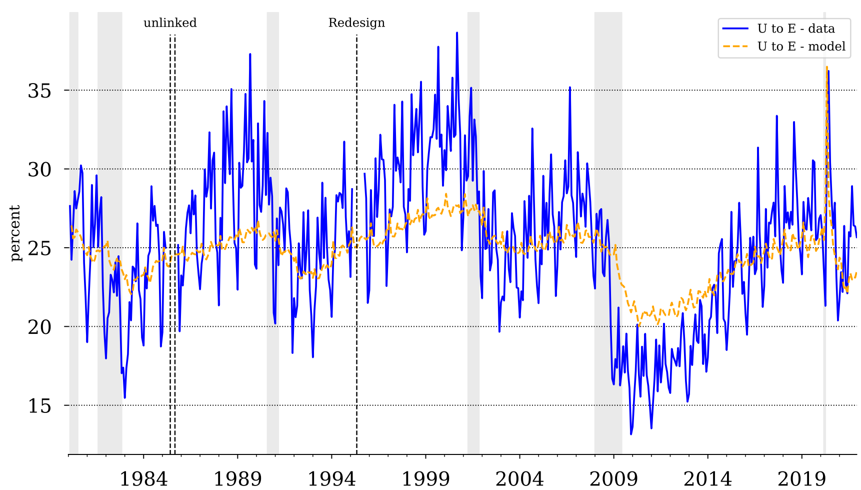
Table B.1: Multi-month transition probabilities in data and models

| spec<br>periods<br>flow | data<br>1-month | model<br>1-month | FOM<br>1-month | data<br>12-month | model<br>12-month | FOM<br>12-month |
|---|---|---|---|---|---|---|
| E to E | 95.61 | 95.40 | 95.74 | 90.09 | 87.06 | 72.21 |
| E to N | 2.95 | 2.97 | 2.74 | 7.26 | 8.90 | 23.97 |
| E to U | 1.43 | 1.63 | 1.52 | 2.65 | 4.04 | 3.82 |
| N to E | 4.55 | 5.02 | 4.45 | 10.76 | 15.54 | 40.65 |
| N to N | 92.96 | 91.99 | 92.88 | 86.92 | 80.79 | 54.57 |
| N to U | 2.49 | 2.99 | 2.67 | 2.32 | 3.67 | 4.78 |
| U to E | 25.26 | 25.27 | 24.35 | 49.99 | 57.07 | 56.59 |
| U to N | 22.69 | 22.48 | 21.25 | 27.90 | 29.61 | 39.03 |
| U to U | 52.06 | 52.25 | 54.40 | 22.11 | 13.33 | 4.38 |

Source: Current Population Survey and authors' calculations.
Notes: Average 1-month and 12-month transition probabilities in data, Dual Labor Market model, and First-Order Markov model.




(a) Time-varying transition and emission probabilities (benchmark)

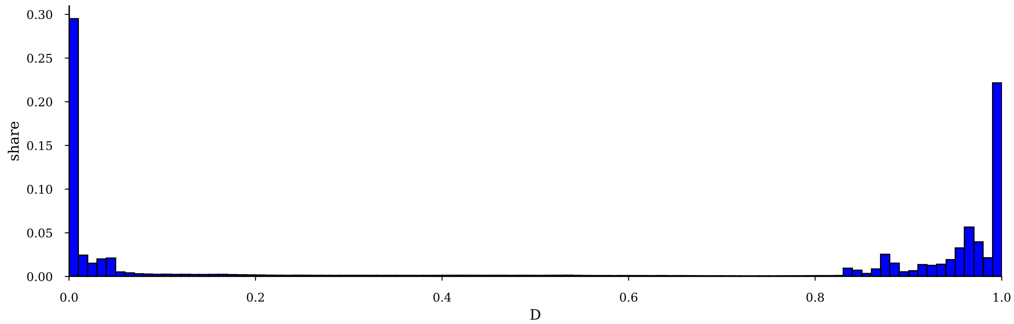



(b) Constant transition and emission probabilities

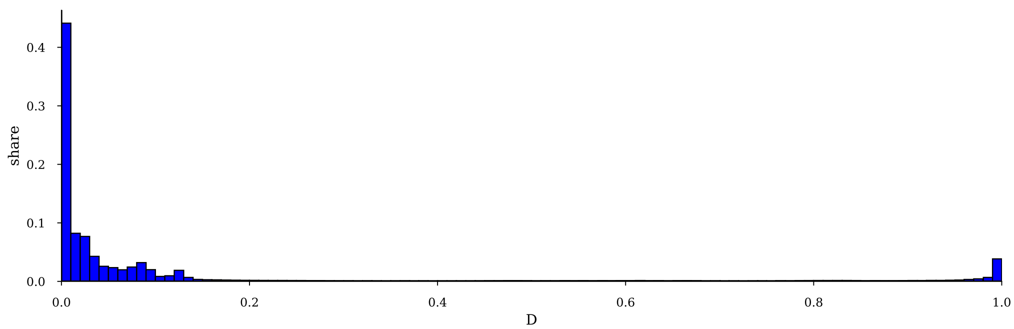
Figure B.1: Monthly actual and estimated job-finding rates ($EU$ probability).

*Source:* CPS and authors' calculations.

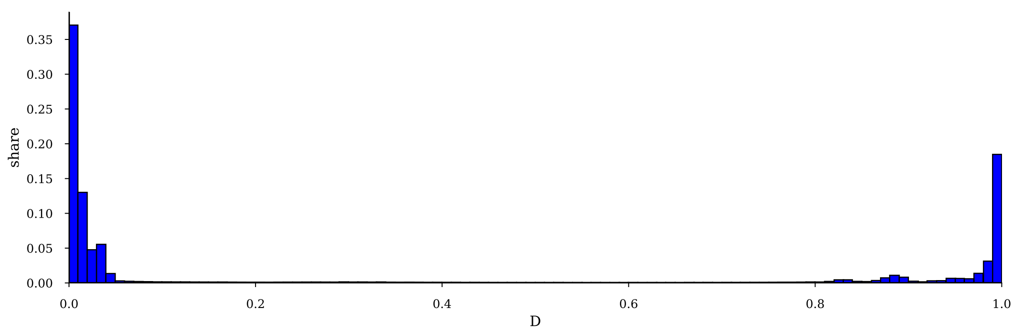
*Notes:* Model with constant transition and emission probabilities allows for a change in the emission probabilities in 1994 to take into account the 1994 CPS redesign.




(a) Primary




(b) Secondary




(c) Tertiary

Figure B.2: Distribution of posterior probabilities by market segment (1980-2021)

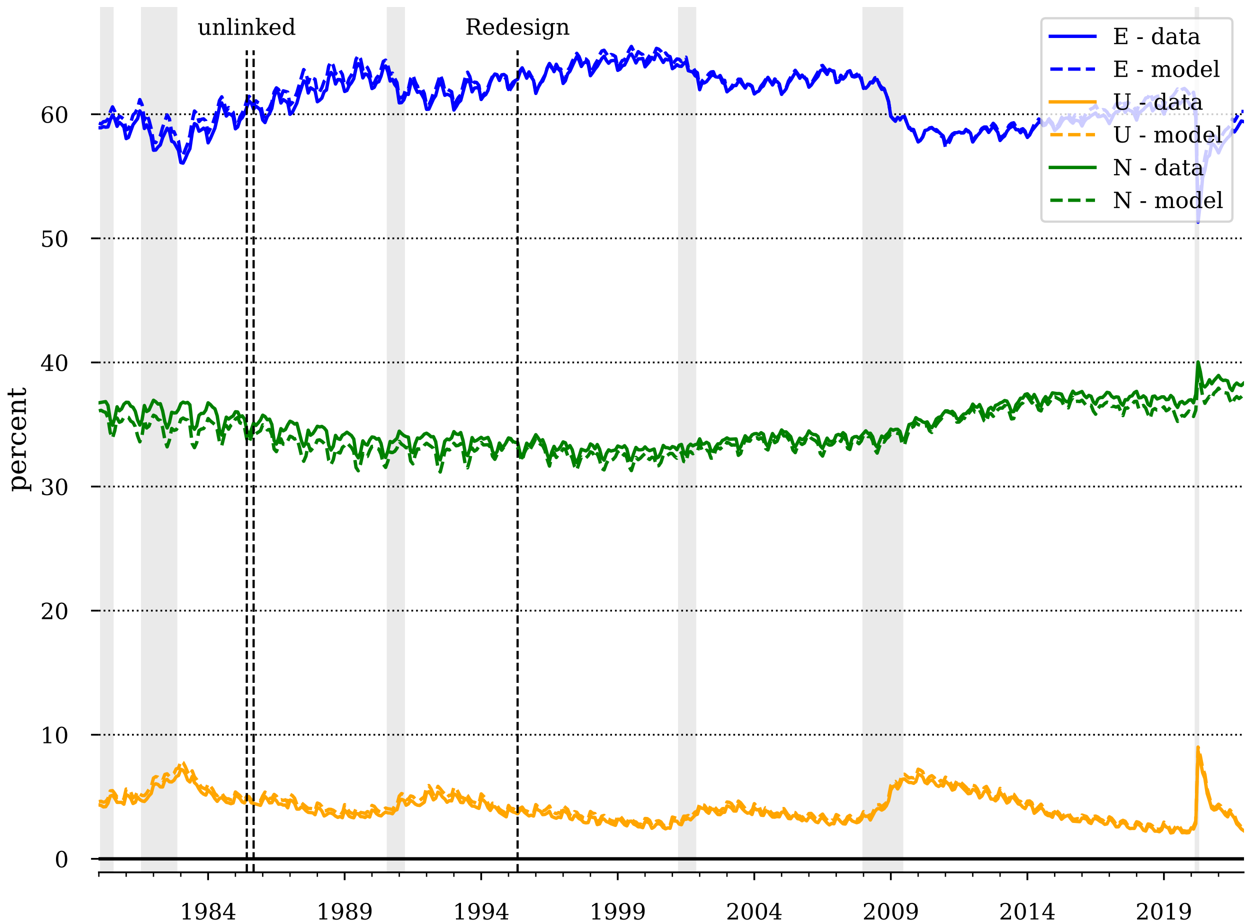
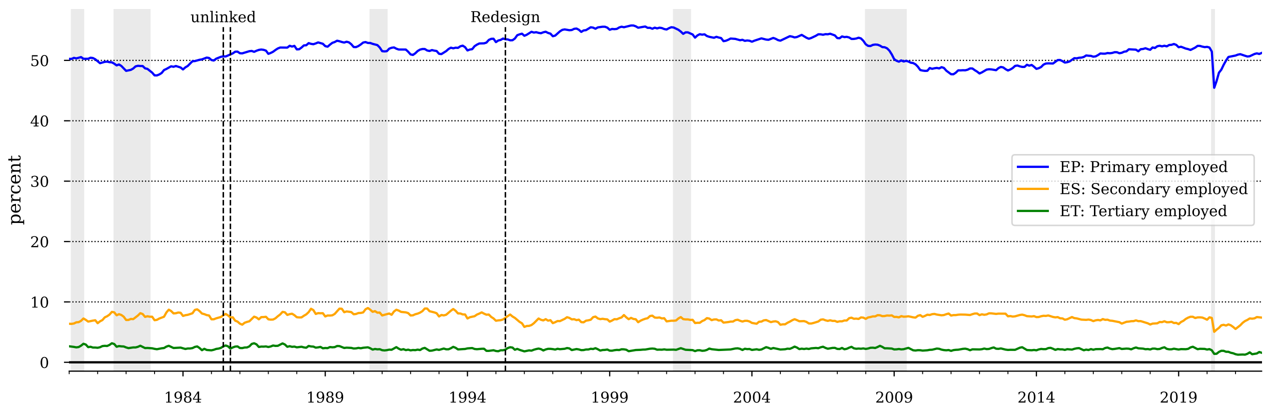


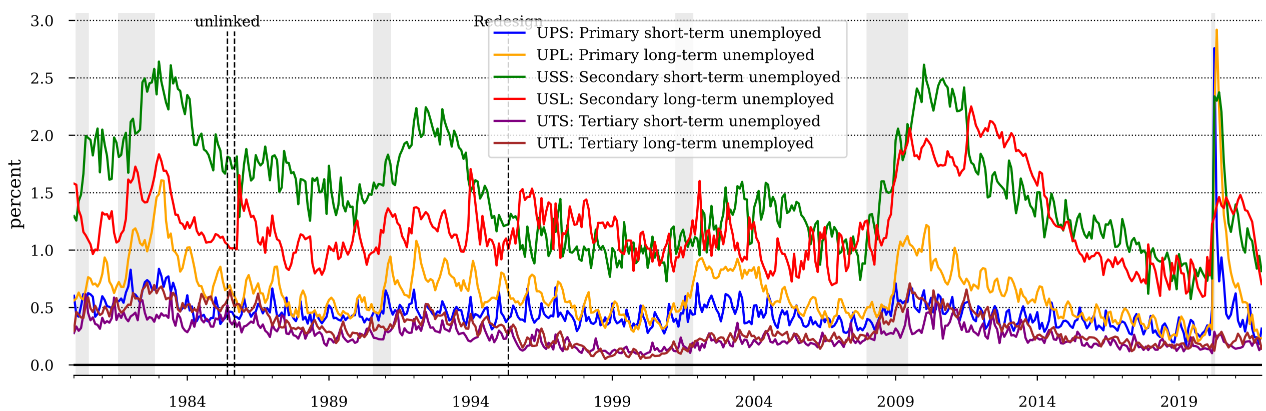

Figure B.3: Actual and fitted shares of population in $E$, $U$, and $N$.

*Source:* CPS and authors' calculations.

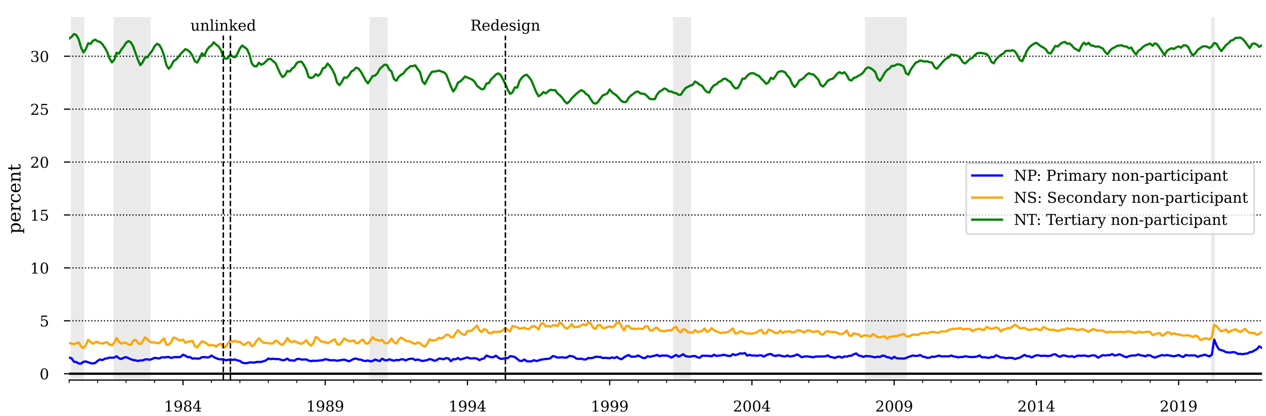
*Notes:* Data and model do not match because of use of 12-month weights in model and imputation of missing values in model estimates.




(a) Employed



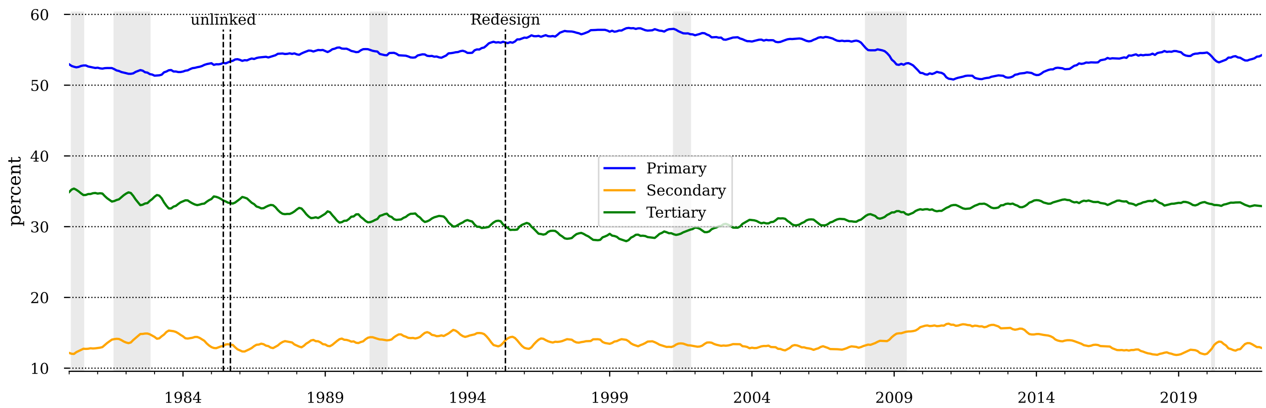

(b) Unemployed



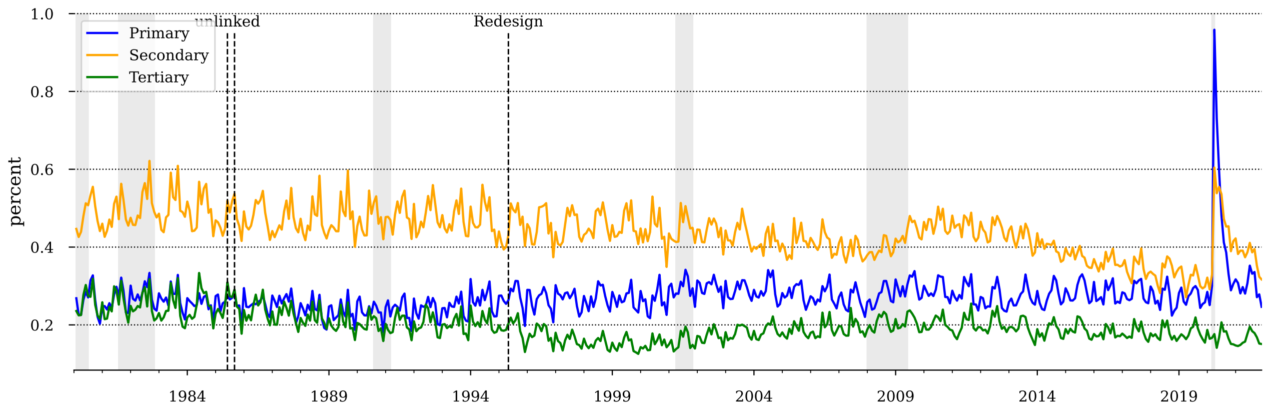

(c) Non-participant

Figure B.4: Estimated shares of population in hidden states.

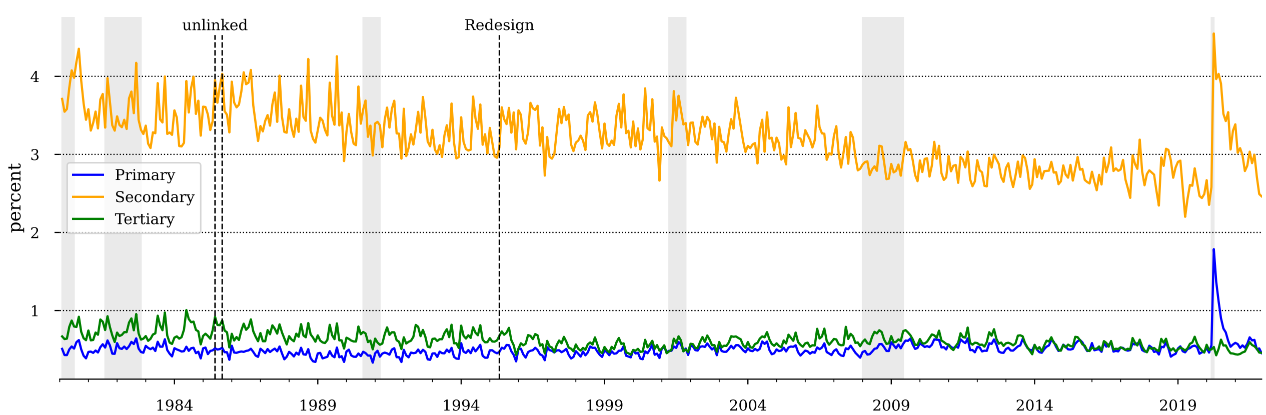
*Source:* CPS and authors' calculations.
*Notes:* "unlinked" is period where household identifiers in CPS are scrambled and 4-8-4 respondent histories are incomplete and "redesign"is the 1994 for which respondent histories are incomplete and the questions used to construct the emissions are changed.




(a) Share of population



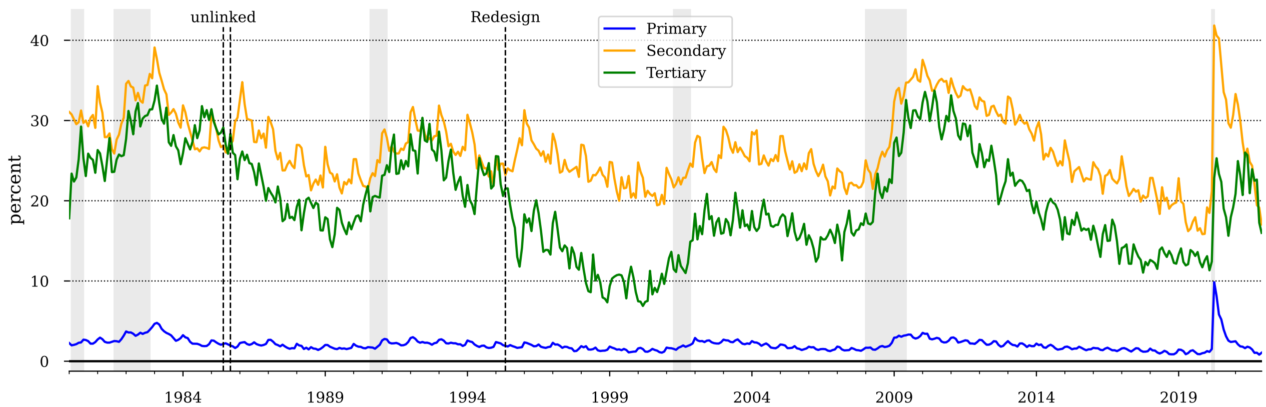

(b) Annualized monthly flows per capita



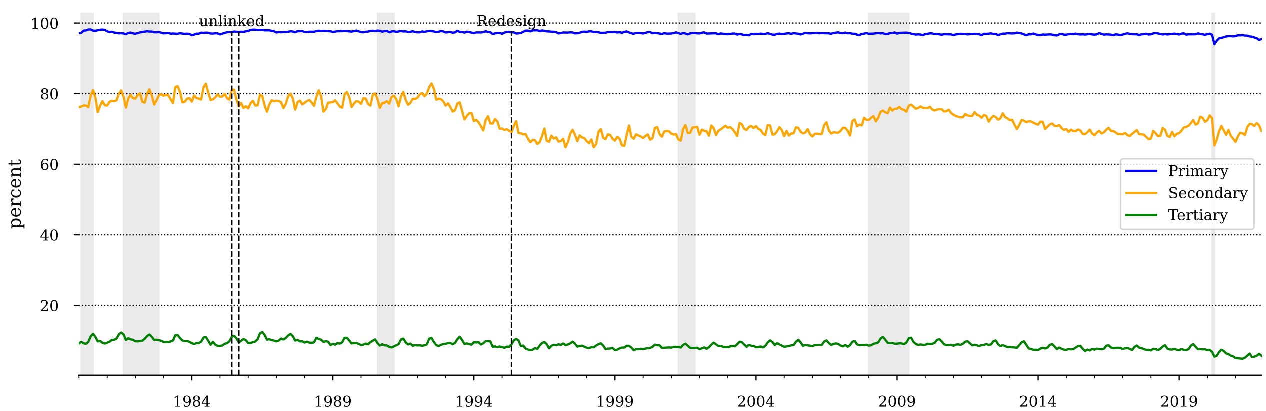

(c) Annualized monthly flows per person in segment

Figure B.5: Shares of population and flows per person by labor market segment.

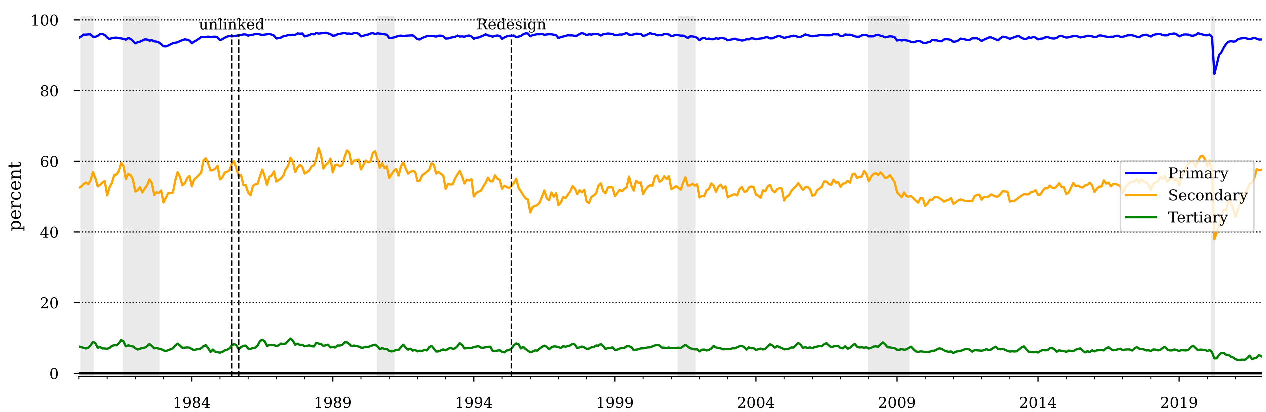
*Source:* CPS and authors' calculations.
*Notes:* "unlinked" is period where household identifiers in CPS are scrambled and 4-8-4 respondent histories are incomplete and "redesign"is the 1994 for which respondent histories are incomplete and the questions used to construct the emissions are changed.




(a) Unemployment rates




(b) Labor force participation rates




(c) Employment-population ratios

Figure B.6: Labor market statistics by tier.

*Source:* CPS and authors' calculations.

*Notes:* "unlinked" is period where household identifiers in CPS are scrambled and 4-8-4 respondent histories are incomplete and "redesign"is the 1994 for which respondent histories are incomplete and the questions used to construct the emissions are changed.

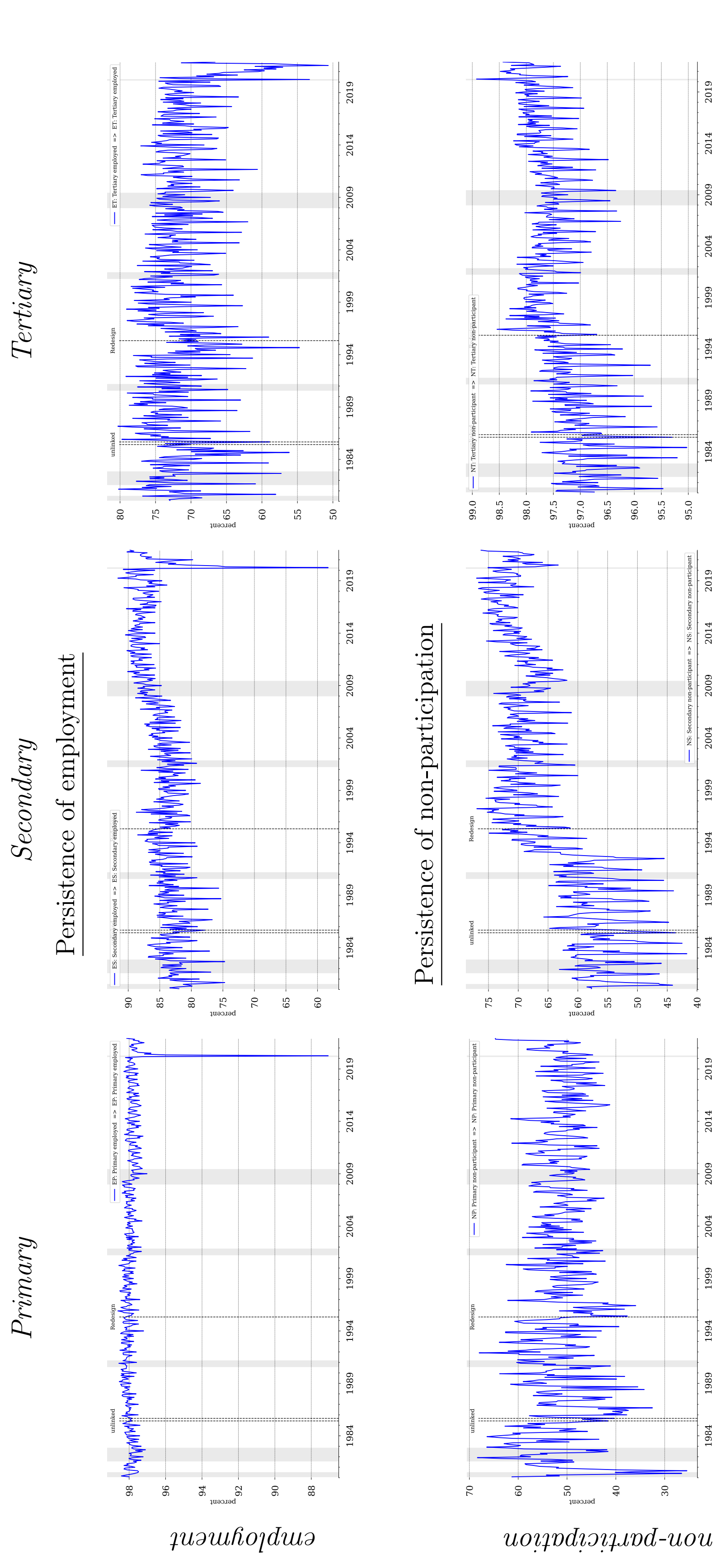



Figure B.7: Persistence of employment and non-participation

*Source:* CPS and authors' calculations. Monthly observations. Not seasonally adjusted. *Note:* Transition probabilities from lont-term to short-term unemployment are restricted to be zero.

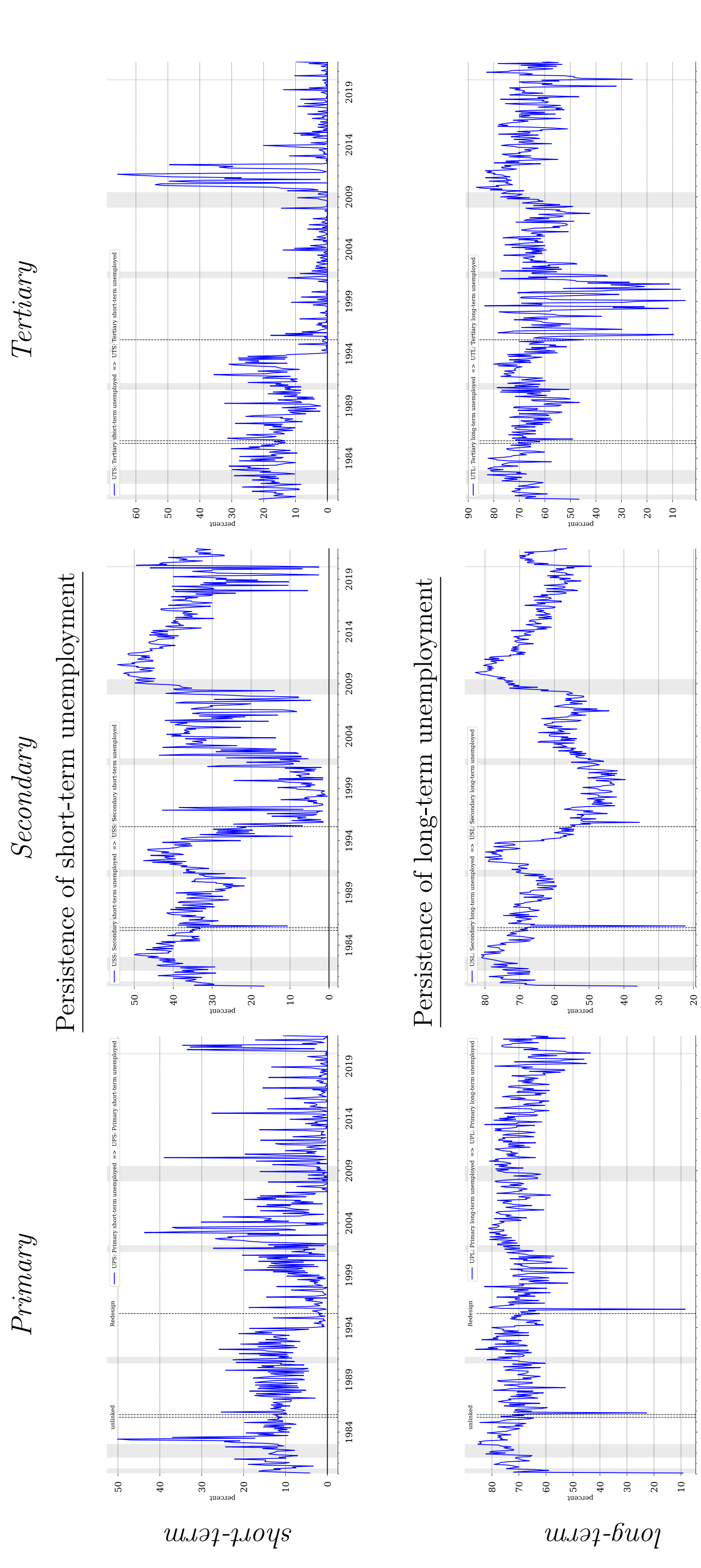



Figure B.8: Persistence of short- and long-term unemployment

*Source:* CPS and authors' calculations. Monthly observations. Not seasonally adjusted. *Note:* Transition probabilities from lont-term to short-term unemployment are restricted to be zero.

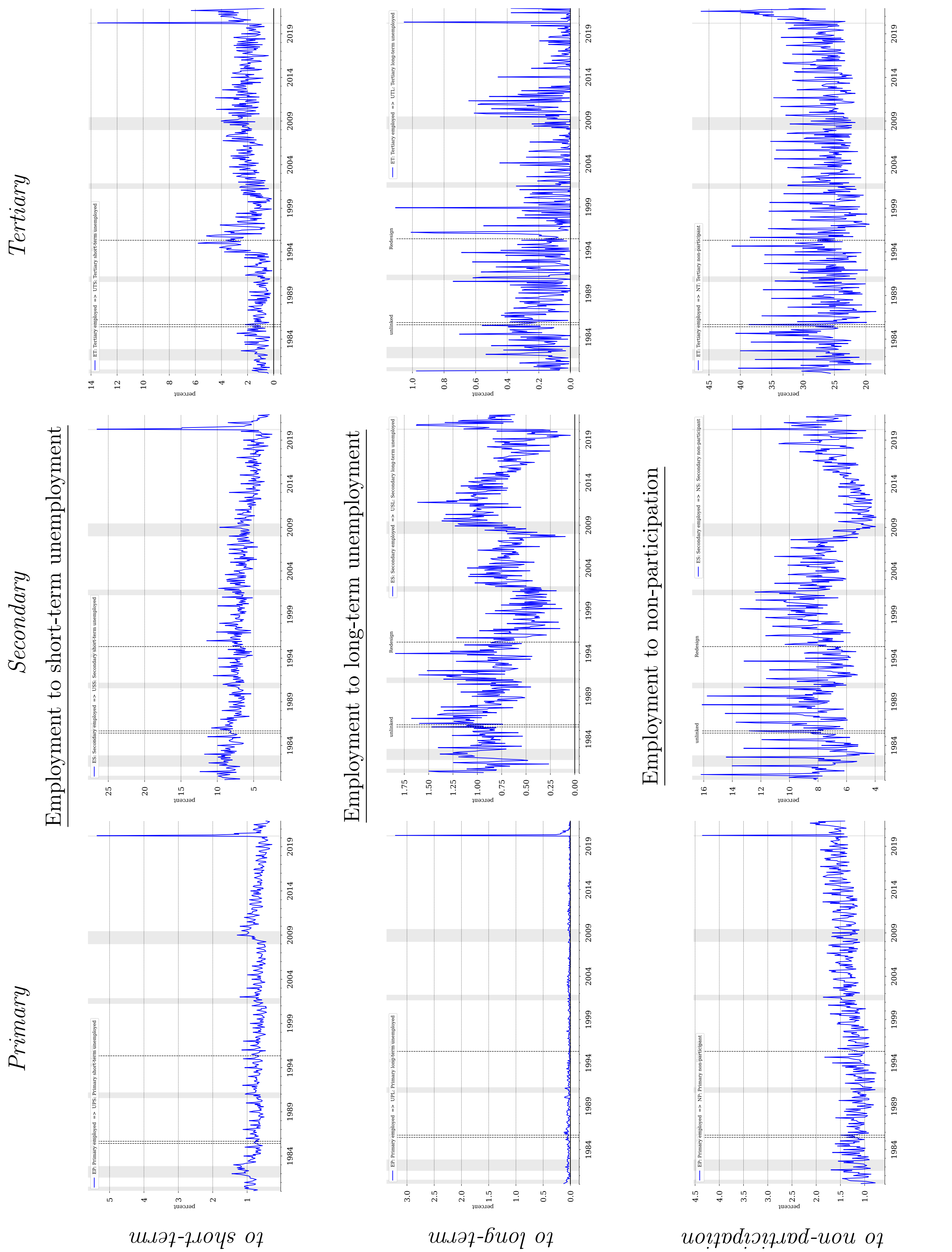



Figure B.9: Estimated transition probabilities from employment

*Source:* CPS and authors' calculations. Monthly observations. Not seasonally adjusted.

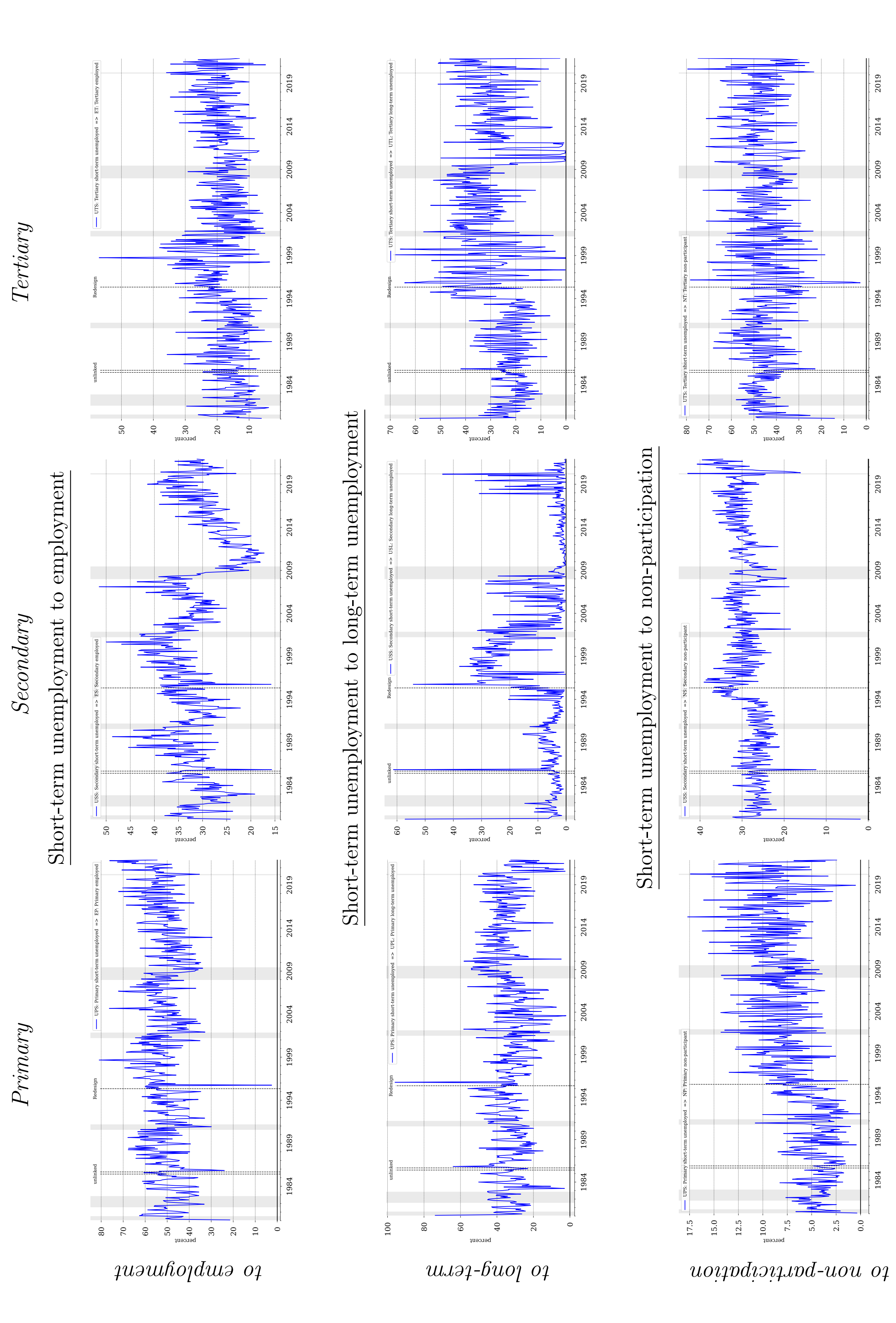



Figure B.10: Estimated transition probabilities from short-term unemployment

*Source:* CPS and authors' calculations. Monthly observations. Not seasonally adjusted.

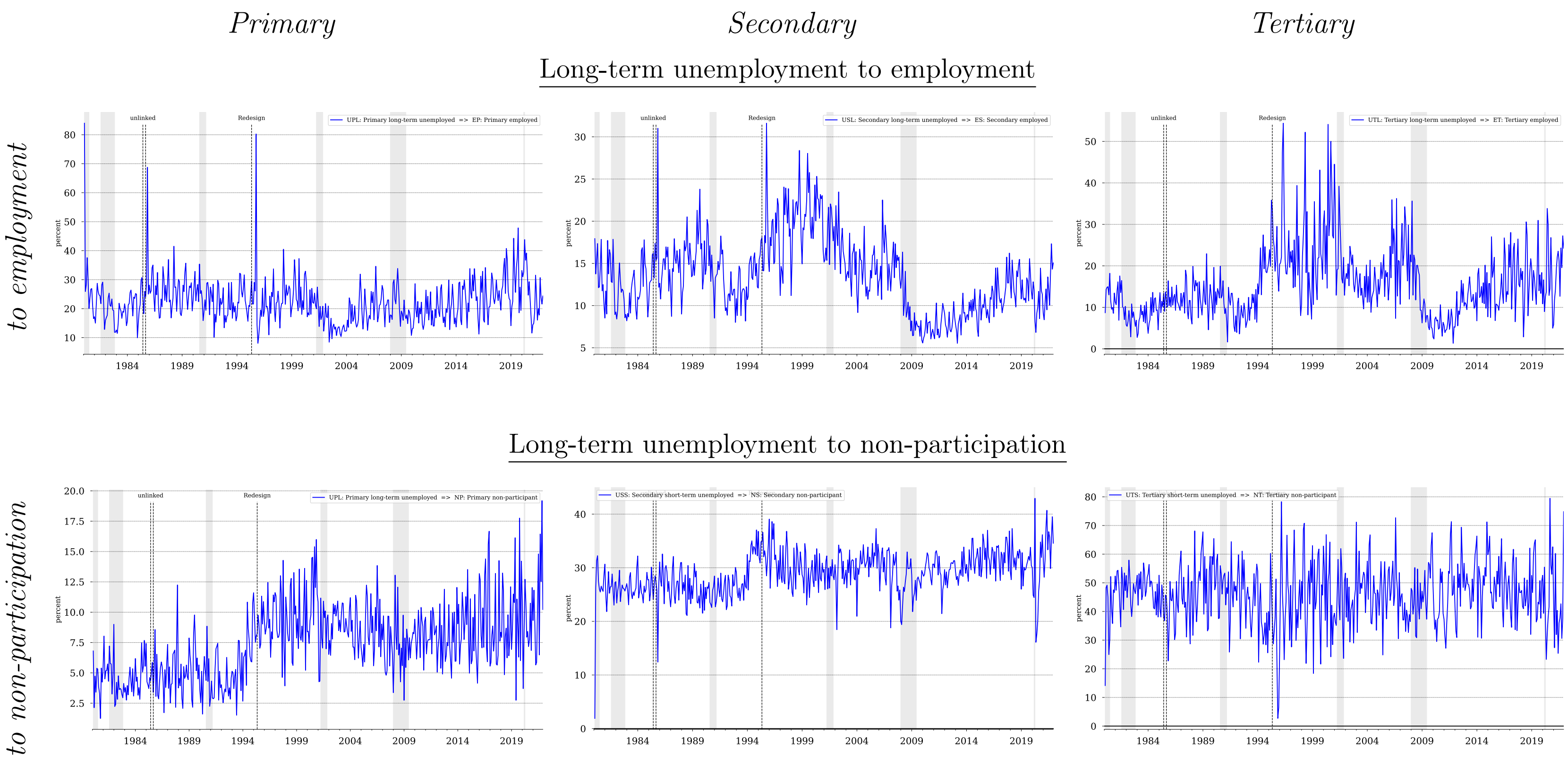



Figure B.11: Estimated transition probabilities from long-term unemployment

*Source:* CPS and authors' calculations. Monthly observations. Not seasonally adjusted. *Note:* Transition probabilities from lont-term to short-term unemployment are restricted to be zero.

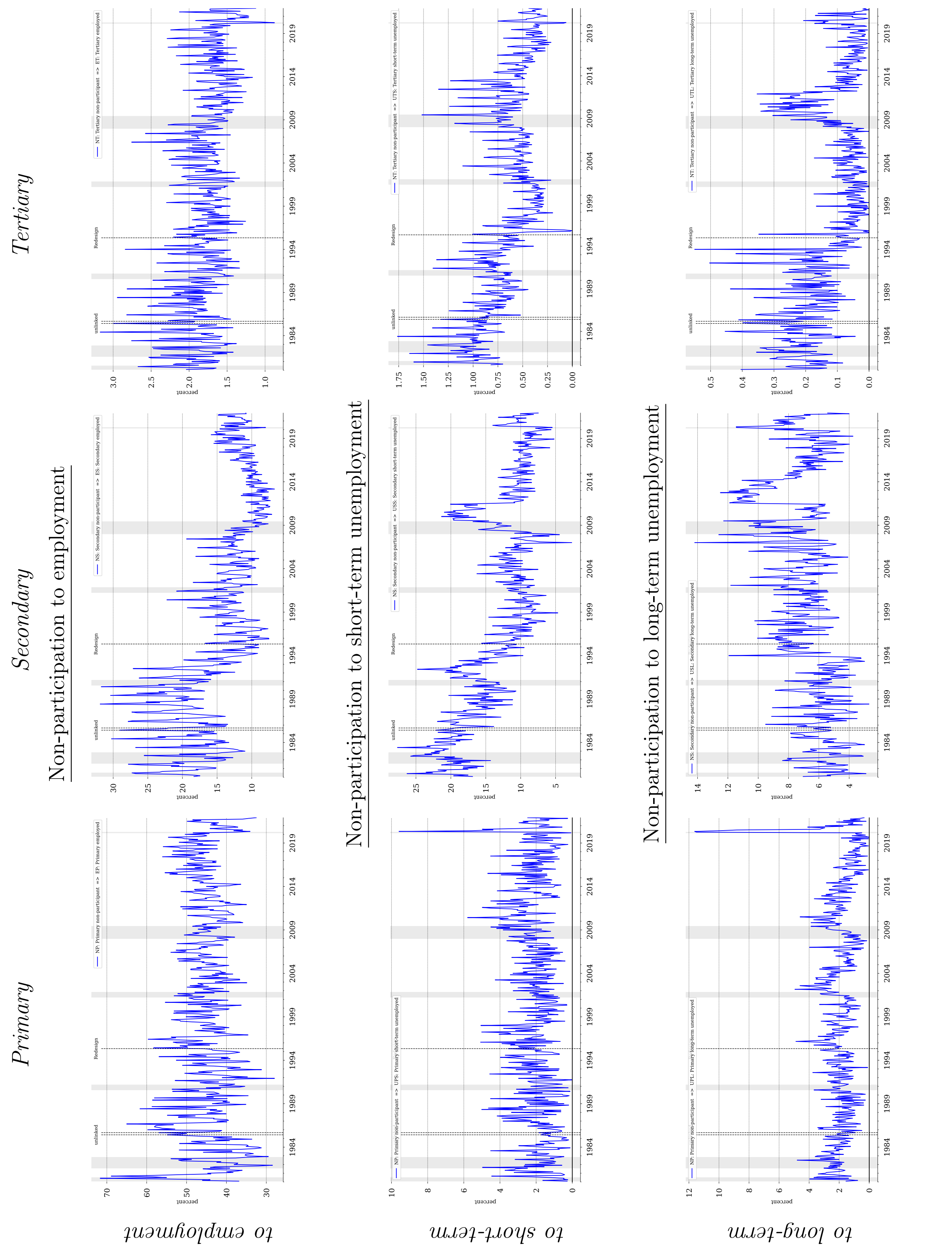
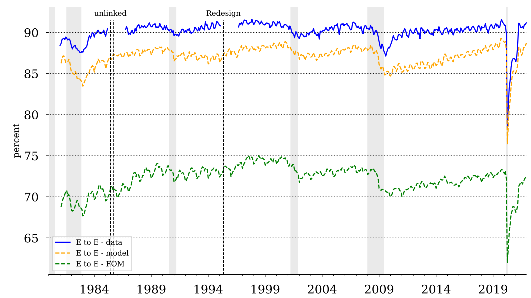
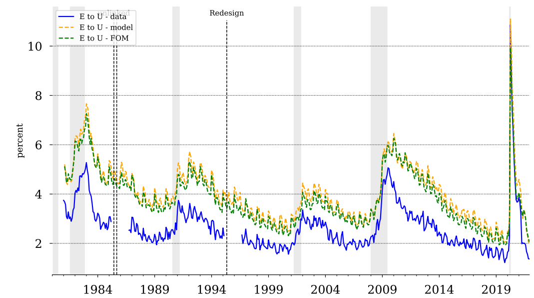



Figure B.12: Estimated transition probabilities from non-participation

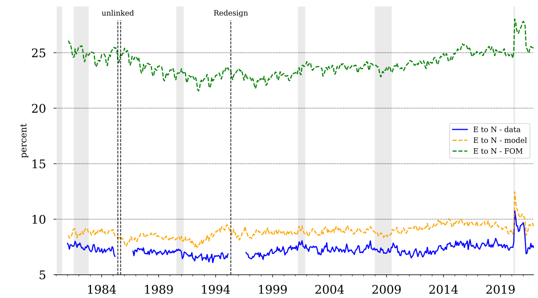
*Source:* CPS and authors' calculations. Monthly observations. Not seasonally adjusted.



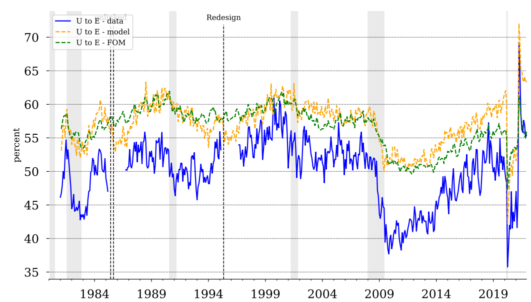

(a) $E \rightarrow E$



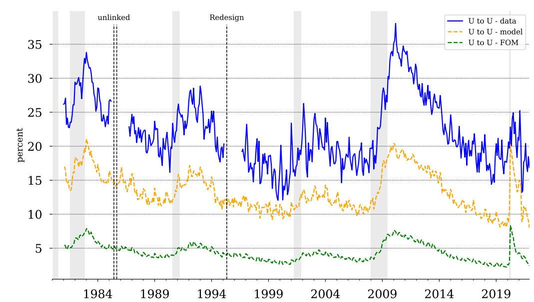

(b) $E \rightarrow U$



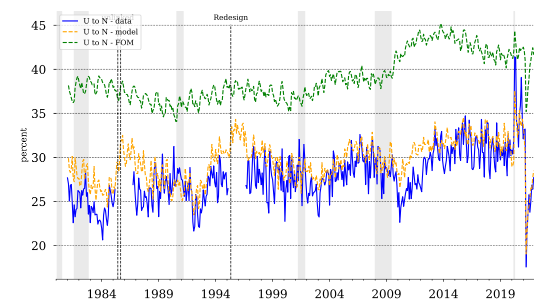

(c) $E \rightarrow N$



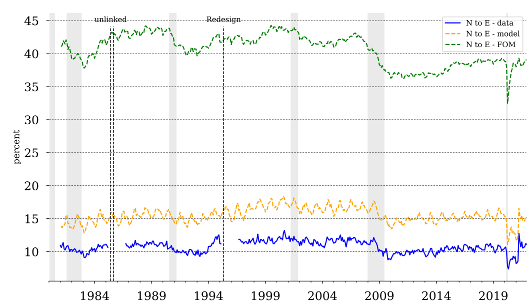

(d) $U \rightarrow E$



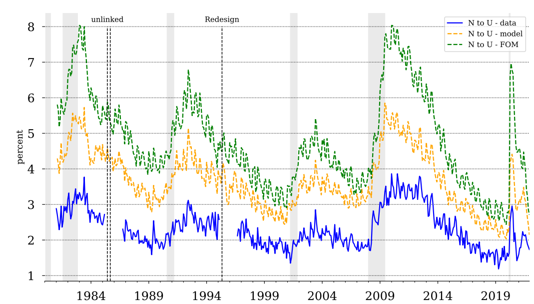

(e) $U \rightarrow U$



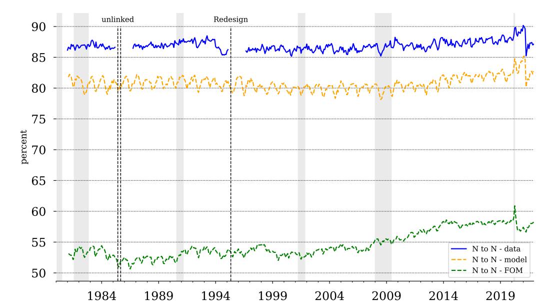

(f) $U \rightarrow N$




(g) $N \rightarrow E$




(h) $N \rightarrow U$




(i) $N \rightarrow N$

Figure B.13: Actual and estimated 12-month transition probabilities between $E$, $U$, and $N$

*Source:* CPS and authors' calculations. Monthly observations. Not seasonally adjusted.

*Notes:* "unlinked" is period where household identifiers in CPS are scrambled and 4-8-4 respondent histories are incomplete and "redesign"is the 1994 for which respondent histories are incomplete and the questions used to construct the emissions are changed. FOM time series are for First-Order Markov model with three states.