A Generalized Model of Misclassification Errors and Labor Force Dynamics

Shuaizhang Feng, Yingyao Hu, and Jiandong Sun*

Abstract

We study the U.S. labor market transitions using a latent variable approach, explicitly modeling the persistent misclassification process and the non-Markovian nature of the underlying true labor force dynamics. A closed-form global identification for misclassification probabilities and labor transition probabilities is established through an eigenvalue-eigenvector decomposition. Contrary to existing studies, our empirical results suggest that the observed data have understated the true mobility in labor force statuses after we account for persistence in both the misclassification errors and the latent true labor force dynamics.

*Feng: Institute for Economic and Social Research, Jinan University, Guangzhou 510632, China (Email: shuaizhang.feng@foxmail.com). Hu: Department of Economics, Johns Hopkins University, Baltimore, MD 21218, United States (Email: yhu@jhu.edu). Sun: Corresponding author, School of Economics and Statistics, Guangzhou University, Guangzhou 510006, China (Email: jiandongsun11@gmail.com). We thank Christopher R. Bollinger, Marc Chan, James J. Heckman, Kai Liu, Robert A. Moffitt, and the participants at the Conference in Honor of Robert A. Moffitt for helpful comments and discussions. Feng acknowledges support from the Key Project of the National Social Science Fund of China (Grant Number: 22AZD079). Sun acknowledges support from the General Project of Humanities and Social Sciences Research of the Ministry of Education of China (Grant Number: 22YJC790108) and the Project of Basic and Applied Basic Research of Guangzhou (Grant Number: 2024A04J3372).

1 Introduction

Gross labor flows describe individuals' labor market experience over the reference period, including how many persons have moved into and out of employment, unemployment or inactivity, rather than just changes in the numbers of persons in these labor force statuses. The U.S. gross labor flows can be tabulated using data matched from the Current Population Surveys (CPS). In each CPS interview, people are asked a set of questions to identify their labor force statuses of either employment (E), unemployment (U) or not-in-the-labor-force (N). However, labor force statuses are misclassified when respondents report inconsistently or interviewers record incorrectly.[1] As a consequence, such misclassification errors lead to inaccuracies in labor force statistics, including both levels of unemployment and changes in labor force statuses.[2]

When considering month-to-month labor flows, misclassified labor force statuses in either period would introduce spurious transitions between two consecutive months. However, the extent to which misclassification errors bias flow statistics depends on the error structure across months (Bound et al. 2001). Lacking empirical evidence, most existing studies rely on the so-called Independent Classification Errors (ICE) assumption, which states that the errors in one month are independent of those in the next (e.g., Abowd and Zellner 1985; Poterba and Summers 1986; Chua and Fuller 1987). The main implication of the ICE assumption is that the observed changes in labor force status would exaggerate the truth due to these spurious transitions, as discussed in Abowd and Zellner (1985).

However, the ICE assumption has been challenged by many researchers. As Singh and Rao (1995) and Bassi and Trivellato (2009) pointed out, misclassification errors might be correlated due to survey design and data collection procedures. A correlated misclassification process might lead the observed labor transitions to display lower mobility than the actual ones, which is opposite to what the ICE assumption would suggest (van de Pol and Langeheine 1997). More recently, Feng et al. (2022) find persistent patterns in the misclassification process using the CPS data, thus empirically rejecting the ICE assumption.[3] Furthermore, existing studies that use a latent variable approach typically assume that the underlying true labor force status follows a first-order Markov process (e.g., Biemer and Bushery 2000; Shibata 2022), which is too strong given the non-Markovian nature of labor force dynamics due to higher-order state dependence, serial correlation, and unobserved heterogeneity discussed in the studies of labor force dynamics (e.g., Hyslop 1999).

[1] Abraham et al. (2013) conduct a validation study using linked CPS data and UI records, and show that 6.4% of the workers who receive wages in the UI records would not report working in the CPS, while 17.6% of those who do not receive wages would report working.

[2] For example, Abowd and Zellner (1985) and Poterba and Summers (1986) correct for misclassification errors in labor flows, while Feng and Hu (2013) focus on how misclassification errors lead to an underestimation of the levels of the U.S. unemployment rate.

[3] Similarly, Keane and Sauer (2010) also identify persistence in misclassification errors in labor force status in their study of married women using the Panel Study of Income Dynamics (PSID).

In this article, we develop a generalized model of the misclassification process and the underlying true labor force dynamics to correct for biases in the flow statistics. Following Feng et al. (2022), we introduce correlation across time in the misclassification process, allowing the current reported status to depend not only on the current true status, but also on the previously-reported status. We further introduce non-stationarity in the misclassification process by allowing misclassification probabilities to be different in two consecutive months, which is a natural relaxation for correcting flow statistics. Regarding the dynamics of the latent true labor force status, we relax the first-order Markov assumption by introducing a lag of reported status into the true labor transition probabilities between any two periods, thus allowing for higher-order state dependence in the underlying true labor force dynamics.[4] Given the joint distribution of observed labor force status and with several reasonable and testable assumptions imposed, we are able to establish a closed-form global identification for both labor transition probabilities and misclassification probabilities, using the eigenvalue-eigenvector decomposition method proposed by Hu (2008). Because our model is proposed to correct for biases in labor flow statistics, identifications and data requirements are different from those in studies that use similar latent variable approaches to investigate other aspects of biases in labor force statistics arising from misclassification errors.[5] To the best of our knowledge, this paper is the first to provide a framework of latent variable with persistence in both the misclassification process and the underlying true labor force dynamics to correct gross labor flows.

To control for heterogeneity, we also condition the misclassification process and the underlying labor force dynamics on observed individual characteristics. Our empirical results show that the previously-reported labor force status does influence the current report even conditional on the current true status, in line with Keane and Sauer (2010) and Feng et al. (2022). The misclassification errors in the fourth interview (rotation group four) are significantly smaller than those in the third interview (rotation group three), implying the non-stationarity of the misclassification process and a possible learning mechanism. Furthermore, we also find higher-order state dependence in the underlying true labor force dynamics, thereby rejecting the first-order Markov assumption commonly used in existing models. Our corrected transition probabilities show higher labor mobility than the ones suggested by the raw data and using existing correction approaches.

The rest of this paper is organized as follows. Section 2 reviews related studies. Section 3 presents a generalized model of the misclassification process and the underlying true labor force dynamics, and discusses our identification assumptions. Section 4 describes the data source, estimation procedure, and simulation results. Section 5 reports empirical results. The last section concludes. Additional

[4] Note that there are alternative ways to model the non-Markovian nature of labor force flows in the literature, such as introducing unobserved heterogeneity (e.g., Shibata 2019; Ahn et al. 2023). We discuss these studies in more detail in Section 2.

[5] For example, Feng and Hu (2013) focus on adjusting labor stock statistics, and Feng et al. (2022) aim to correct for rotation group bias.

results are included in the online Appendix.

2 Literature Review

Many researchers have tried to correct for misclassification errors when estimating gross labor flows. Most earlier studies rely on the ICE assumption, which states that the probability of observing a flow (i, j) between months t and $t+1$ when the true flow is (k, l) is the product of the probability of observing status i in month t when the true status is k and the probability of observing status j in month $t+1$ when the true status is l. That is,

$$\Pr\left(S_t = i, S_{t+1} = j | S_t^* = k, S_{t+1}^* = l\right) = \Pr\left(S_t = i | S_t^* = k\right) \Pr\left(S_{t+1} = j | S_{t+1}^* = l\right),$$

where $S_\tau, S_\tau^* \in \{E, U, N\}$ for $\tau \in \{t, t+1\}$ are the observed and the true statuses in month τ, respectively. The ICE assumption implies that: (i) classification errors between two months are independent conditional on the true statuses, i.e., serial independence; (ii) classification errors only depend on the current true statuses, i.e., local independence or transition independence (Meyer 1988; Singh and Rao 1995).

Under the ICE assumption with three labor force statuses, the misclassification process can be summarized as a simple 3-by-3 misclassification matrix, which maps the true statuses into the observed ones.[6] To identify such a misclassification matrix, early studies use external information, such as reinterview surveys. Some of them treat the responses from the reconciled reinterview surveys as the "truth" (e.g., Abowd and Zellner 1985; Poterba and Summers 1986; Magnac and Visser 1999). But the reinterview surveys also suffer from misclassification errors due to many practical limitations, and may contain even more errors than the original sample (Biemer and Forsman 1992; Sinclair and Gastwirth 1996). Other studies impose alternative but also strong assumptions to achieve model identification (e.g., Chua and Fuller 1987; Sinclair and Gastwirth 1996). For example, Sinclair and Gastwirth (1996) assume that the misclassification probabilities are the same for different subsamples. Such restrictive assumptions, however, contradict to the findings of heterogeneous misclassification probabilities across demographic groups (e.g., Poterba and Summers 1986; Feng and Hu 2013).

Utilizing the panel structure of labor force surveys, more recent studies take a latent variable

[6]Some other studies also use a so-called "DeNUNified" method, which directly change some labor force sequences considered as spurious. Specifically, this procedure recodes UNU and NUN transitions as UUU and NNN transitions, such as Rothstein (2011), Elsby et al. (2015), and Farber and Valletta (2015). However, other studies argue that these short-term labor transition reversals cannot be totally explained as misclassification errors, and the "DeNUNified" method is not a reasonable approach to deal with these reversals (e.g., Kudlyak and Lange 2018; Hall and Kudlyak 2019; Gregory et al. 2021). Ahn and Hamilton (2022) modify this method and only re-classify UNU as UUU if the final U reports 5-week or longer unemployment duration, as they find that the job-finding probabilities reported by people who make NU and UU transitions with 5-week or longer unemployment duration are essentially the same.

approach to model the underlying true labor force dynamics and the misclassification process simultaneously, such as Biemer and Bushery (2000), Bassi and Trivellato (2009), Feng and Hu (2013), Shibata (2022), and Feng et al. (2022).[7] A few of these studies challenge the ICE assumption imposed on the misclassification process. Bassi and Trivellato (2009) argue that the ICE assumption seems unrealistic if panel data is collected retrospectively. They then consider a scenario when misclassification errors are correlated across periods. Specifically, they assume that the reported status in period t depends on the true statuses in both periods t and $t+1$, as they use retrospective dataset and information in period t is collected in period $t+1$. Differently, Feng et al. (2022) assume that the reported status in period t depends not only on the current true status, but also on the reported status in period $t-1$, which is more reasonable in the CPS context. Similarly, Keane and Sauer (2010) also find persistent misreporting behavior in the labor force status of married women using the PSID data.

In terms of modeling the labor force dynamics, researchers have developed structural models where workers maximize expected payoffs by considering their decisions of labor force participation and job acceptance, such as Toikka (1976) and Flinn and Heckman (1982) among others. Dynamic discrete choices models, including Markov models, Pólya schemes and so on, are then established to approximate the decision rules derived from structural models (see e.g., Heckman 1981b; Burdett et al. 1984). However, these studies do not explicitly consider measurement errors in the labor force statuses. On the other hand, latent variable models that incorporate measurement errors assume the underlying true labor force status to follow a first-order Markov process (e.g., Biemer and Bushery 2000; Bassi and Trivellato 2009; Shibata 2022), which might be too strong, as many factors could cause labor force flows to exhibit non-Markovian characteristics, such as higher-order state dependence, serial correlation, and unobservable heterogeneity (e.g., Heckman and Willis 1977; Heckman 1981a; Hyslop 1999; Magnac 2000; Prowse 2012).[8]

A few recent studies have tried to relax the first-order Markov assumption by making use of the eight drop-out periods in the CPS. For example, Feng and Hu (2013) assume that the true status nine months ago has no predictive power over the underlying true labor force dynamics across two consecutive months, while Feng et al. (2022) assume that conditional on the current true status and the reported status in the previous month, earlier true or reported statuses do not influence the underlying true labor force dynamics across nine months. However, Feng and Hu (2013) and Feng et al. (2022) study biases in labor stock statistics, and their assumptions cannot be directly used to correct labor force flows.

Our paper is also related to recent studies that model the non-Markovian nature of labor force

[7] Please refer to Hu (2017) for more details about the latent variable approach.

[8] Also see evidence on duration dependence (e.g., Burdett et al. 1985; van den Berg and van Ours 1996; Kroft et al. 2013; Eriksson and Rooth 2014; Krueger et al. 2014; Kroft et al. 2016), which is a type of state dependence. For detailed distinction, we refer to Heckman and Borjas (1980).

flows from the perspective of unobservable heterogeneity, such as Shibata (2019) and Ahn et al. (2023). Specifically, both studies assume that there are more latent statuses than the three observed labor force statuses (i.e., E, U, and N). Based on this, they assume that their extended latent statuses follow the first-order Markovian process. This is a reasonable way of modeling the non-Markovian nature of labor force flows, which can also be supported by rich evidence on unobserved heterogeneity (e.g., Ahn and Hamilton 2020). However, Ahn et al. (2023) assume there are no misclassification errors, which might be too strong given the evidence on the presence of misclassification errors (e.g., Abowd and Zellner 1985; Poterba and Summers 1986; Abraham et al. 2013). Shibata (2019) imposes the local independence assumption.

3 Model and Identification

3.1 A model of misclassification process and true labor force dynamics

In this subsection, we model the misclassification process and the underlying true labor force dynamics based on the 4-8-4 rotating structure of the CPS.[9] Suppose that for each individual in a random sample, we observe the reported statuses for two spells of four consecutive months as follows:

$$\left\{S_{t-2}^{1}, S_{t-1}^{2}, S_{t}^{3}, S_{t+1}^{4}, S_{t+10}^{5}, S_{t+11}^{6}, S_{t+12}^{7}, S_{t+13}^{8}\right\},$$

where the superscript is month-in-sample (MIS) and the subscript is calendar month. Between the fourth and fifth interviews, there are eight drop-out months, during which we have no information. For simplicity, we omit the superscript hereafter. The reported status is defined as follows:

$$S_t = \begin{cases} 1 & \text{employed } (E) \\ 2 & \text{unemployed } (U) \\ 3 & \text{not-in-the-labor-force } (N) \end{cases}.$$

The latent true status S_t^* shares the same support as the reported status.

Let $\Pr(\cdot)$ stand for the probability distribution function of its arguments. We follow Feng et al. (2022) to outline the assumption with respect to the reporting behavior as follows:

Assumption 1. *Conditional on observed characteristics $\boldsymbol{X}$, the reported status in the current month (S_t) only depends on the true status in the current month (S_t^*) and the reported status in the previous month (S_{t-1}), i.e.,*

$$\Pr\left(S_t|S_t^*, \{S_\tau^*, S_\tau\}_{\tau\le t-1}, \boldsymbol{X}\right) = \Pr\left(S_t|S_t^*, S_{t-1}, \boldsymbol{X}\right).$$

[9] See Section A in the online Appendix for an illustration of the 4-8-4 rotating structure in the CPS.

Assumption 1 allows for the correlation between the reported statuses across two consecutive months even conditional on the current true status, which is much weaker than the widely-used local independence assumption. This assumption can be supported by studies on panel conditioning, that is, repeated participation in a panel survey might change respondents' reporting behavior and other aspects (e.g., Halpern-Manners and Warren 2012). Note that respondents are not interviewed for those drop-out periods, implying that the reported status may only depend on the true status for incoming rotation groups (i.e., rotation groups one and five). That is, for $i \in \{t-2, t+10\}$,

$$\Pr\left(S_i | S_i^*, \{S_\tau^*, S_\tau\}_{\tau \leq i-1}, \mathbf{X}\right) = \Pr\left(S_i | S_i^*, \mathbf{X}\right).$$

Having assumed the conditional independence of the misclassification process, we next impose the following assumption on the underlying true labor force dynamics:

Assumption 2. *Conditional on observed characteristics $\boldsymbol{X}$, the true status in the current month (S_t^*) and the reported status in the previous month (S_{t-1}), the true or reported statuses in other months have no predictive power on the true status p months later (S_{t+p}^*). That is, for $p \geq 1$,*

$$\Pr\left(S_{t+p}^* | \{S_\tau^*, S_\tau\}_{\tau \leq t}, \boldsymbol{X}\right) = \Pr\left(S_{t+p}^* | S_t^*, S_{t-1}, \boldsymbol{X}\right).$$

Assumption 2 incorporates the non-Markovian nature of the true labor force dynamics by including a lag of reported status in the true labor force transition across periods t and $t+p$, meaning that historical labor market experiences could affect future labor market outcomes, which is also in line with Kudlyak and Lange (2018).[10] This assumption can be supported by evidence on higher-order state dependence and duration dependence of unemployment hazards (e.g., Kroft et al. 2013; Ahn and Hamilton 2020).[11] Technically, it is much weaker than the first-order Markov assumption that the whole process of $\{S_t, S_t^*\}$ is Markovian and $\Pr\left(S_{t+1}^* | S_t^*, S_t\right) = \Pr\left(S_{t+1}^* | S_t^*\right)$, such as in Biemer and Bushery (2000) and Shibata (2022).

[10] Although our Assumptions 1 and 2 relax previous models, both assumptions can be extended to include more lags of either reported or true status. However, doing so entails additional costs in terms of more restrictive assumptions in other dimensions and much larger standard errors. We investigate these possibilities in Section 4.3 and Section 5.3.

[11] More broadly, higher-order state dependence and duration dependence are also investigated among the employed people and nonparticipants. For example, Burdett et al. (1985) model the duration dependence in employment and find that a declining hazard rate in the transition out of employment is a significant feature of the data. Kudlyak and Lange (2018) find that those nonparticipants with recent unemployment have a much higher job finding rate than those who have been out-of-labor-force for three consecutive months.

3.2 A closed-form global identification

Under Assumptions 1 and 2, we derive the following joint distribution:

$$\begin{aligned} & \Pr\left(S_{t+10}, S_{t+1}, S_t, S_{t-1}, S_{t-2} | \mathbf{X}\right) \\ = & \sum_{S_{t+1}^*} \Pr\left(S_{t+10} | S_{t+1}^*, S_t, \mathbf{X}\right) \Pr\left(S_{t+1} | S_{t+1}^*, S_t, \mathbf{X}\right) \Pr\left(S_{t+1}^*, S_t, S_{t-1}, S_{t-2} | \mathbf{X}\right). \end{aligned} \tag{1}$$

This means that, if S_t and S_{t-1} are fixed, we may apply the identification strategy in Hu (2008) to identify the unknown conditional distributions on the right-hand side of Equation (1). Integrating out S_{t+10}, we have

$$\Pr\left(S_{t+1}, S_t, S_{t-1}, S_{t-2} | \mathbf{X}\right) = \sum_{S_{t+1}^*} \Pr\left(S_{t+1} | S_{t+1}^*, S_t, \mathbf{X}\right) \Pr\left(S_{t+1}^*, S_t, S_t, S_{t-2} | \mathbf{X}\right). \tag{2}$$

Next, we specify the main procedures of identifying the misclassification probabilities and the labor force transition probabilities. For any given sub-population with $\mathbf{X} = \mathbf{x}$, we start with $S_{t+10} = 1$, $S_t = s_t$ and $S_{t-1} = s_{t-1}$.[12] As shown in Hu (2008), Equations (1) and (2) imply the following two matrix equations:

$$M_{1,S_{t+1},s_t,s_{t-1},S_{t-2}|\mathbf{x}} = M_{S_{t+1}|S_{t+1}^*,s_t,\mathbf{x}} D_{1|S_{t+1}^*,s_t,\mathbf{x}} M_{S_{t+1}^*,s_t,s_{t-1},S_{t-2}|\mathbf{x}}, \tag{3}$$

$$M_{S_{t+1},s_t,s_{t-1},S_{t-2}|\mathbf{x}} = M_{S_{t+1}|S_{t+1}^*,s_t,\mathbf{x}} M_{S_{t+1}^*,s_t,s_{t-1},S_{t-2}|\mathbf{x}}, \tag{4}$$

where

$$\begin{aligned} M_{1,S_{t+1},s_t,s_{t-1},S_{t-2}|\mathbf{x}} &\equiv \left[\Pr\left(S_{t+10} = 1, S_{t+1} = i, S_t = s_t, S_{t-1} = s_{t-1}, S_{t-2} = j | \mathbf{x}\right)\right]_{i,j}, \\ M_{S_{t+1},s_t,s_{t-1},S_{t-2}|\mathbf{x}} &\equiv \left[\Pr\left(S_{t+1} = i, S_t = s_t, S_{t-1} = s_{t-1}, S_{t-2} = j | \mathbf{x}\right)\right]_{i,j}, \\ D_{1|S_{t+1}^*,s_t,\mathbf{x}} &\equiv Diag\left[\Pr\left(S_{t+10} = 1 | S_{t+1}^* = j, S_t = s_t, \mathbf{x}\right)\right]_j, \\ M_{S_{t+1}|S_{t+1}^*,s_t,\mathbf{x}} &\equiv \left[\Pr\left(S_{t+1} = i | S_{t+1}^* = j, S_t = s_t, \mathbf{x}\right)\right]_{i,j}, \\ M_{S_{t+1}^*,s_t,s_{t-1},S_{t-2}|\mathbf{x}} &\equiv \left[\Pr\left(S_{t+1}^* = i, S_t = s_t, S_{t-1} = s_{t-1}, S_{t-2} = j | \mathbf{x}\right)\right]_{i,j}. \end{aligned}$$

The misclassification probabilities matrix, i.e., $M_{S_{t+1}|S_{t+1}^*,s_t}$, can be identified from an eigenvalue-eigenvector decomposition if the following two assumptions hold.

Assumption 3. *For each s_t and s_{t-1}, $M_{S_{t+1},s_t,s_{t-1},S_{t-2}|\boldsymbol{x}}$ has full rank.*

Assumption 4. $\Pr\left(S_{t+10} = 1 | S_{t+1}^* = i, S_t = s_t, \boldsymbol{x}\right) \neq \Pr\left(S_{t+10} = 1 | S_{t+1}^* = j, S_t = s_t, \boldsymbol{x}\right)$ *for $i \neq j$.*

[12] The following identification argument holds for each value of S_{t+10}, implying that the model is over-identified.

Assumption 3 is imposed on the observed probabilities, meaning that it can be tested using the CPS data directly. In Section E in the online Appendix, we use bootstrapping to show that the determinants of the observed matrices are significantly different from zero, which implies that these matrices are invertible. We then can eliminate $M_{S^*_{t+1},s_t,s_{t-1},S_{t-2}|\mathbf{x}}$ in Equations (3) and (4), and derive the following equation:

$$M_{1,S_{t+1},s_t,s_{t-1},S_{t-2}|\mathbf{x}}M^{-1}_{S_{t+1},s_t,s_{t-1},S_{t-2}|\mathbf{x}} = M_{S_{t+1}|S^*_{t+1},s_t,\mathbf{x}}D_{1|S^*_{t+1},s_t,\mathbf{x}}M^{-1}_{S_{t+1}|S^*_{t+1},s_t,\mathbf{x}}, \tag{5}$$

which shows that the observed matrix on the left-hand side of Equation (5) has an eigenvalue-eigenvector decomposition on the right-hand side. Assumption 4 ensures that the eigenvalues are distinctive and the eigenvectors can be uniquely identified, which is also testable using the CPS data.[13]

In order to determine the ordering of the eigenvectors, we impose the following assumption:

Assumption 5. *Given $S_\tau = k$, (i) $\Pr\left(S_{\tau+1} = k|S^*_{\tau+1} = k, S_\tau = k, \boldsymbol{x}\right)$ is the largest element in row k; (ii) for $i \neq j \neq k$, $\Pr\left(S_{\tau+1} = i|S^*_{\tau+1} = j, S_\tau = k, \boldsymbol{x}\right)$ is the smallest element in column j.*

Assumption 5 implies that given the previously-reported status, if the current true status is the same as the previously-reported status, individuals are always more likely to report that status than if the true status is otherwise. Further, if the current true status is different from the previously-reported status, then the least possible choice to report would be the status other than the current true status or the previously-reported status. Assumption 5 is an intuitive extension of the standard assumption in the literature when the reported status is assumed to only depend on the latent true status, and people are more likely to report the truth than otherwise. In our extended framework, both latent true status and previously-reported status could matter.

Once we have identified the misclassification probabilities in period $t+1$, i.e., $\Pr\left(S_{t+1}|S^*_{t+1}, S_t, \mathbf{X}\right)$, we can estimate $\Pr\left(S^*_{t+1}, S_t, S_{t-1}, S_{t-2}|\mathbf{X}\right)$ from Equation (2). To further identify the misclassification probabilities in period t and the transition probabilities, i.e., $\Pr\left(S_t|S^*_t, S_{t-1}, \mathbf{X}\right)$ and $\Pr\left(S^*_{t+1}|S^*_t, S_{t-1}, \mathbf{X}\right)$, we again apply the same strategy to the following equation:

$$\Pr\left(S^*_{t+1}, S_t, S_{t-1}, S_{t-2}|\mathbf{X}\right) = \sum_{S^*_t} \Pr\left(S^*_{t+1}|S^*_t, S_{t-1}, \mathbf{X}\right) \Pr\left(S_t|S^*_t, S_{t-1}, \mathbf{X}\right) \Pr\left(S^*_t, S_{t-1}, S_{t-2}|\mathbf{X}\right). \tag{6}$$

To guarantee the eigenvalue-eigenvector decomposition can be proceeded, we make the following two assumptions:

[13] We would not report testing results on this assumption here, as we will estimate the model parametrically, as shown in Section 4.2. Similar tests can be found in Feng et al. (2022).

Assumption 6. *For each s_{t-1}, $M_{S_t^*, s_{t-1}, S_{t-2}|\boldsymbol{x}}$ has full rank.*[14]

Assumption 7. $\Pr\left(S_{t+1}^* = 1 | S_t^* = i, S_{t-1} = s_{t-1}, \boldsymbol{x}\right) \neq \Pr\left(S_{t+1}^* = 1 | S_t^* = j, S_{t-1} = s_{t-1}, \boldsymbol{x}\right)$ *for* $i \neq j$.

Finally, we summarize the closed-form identification of the misclassification probabilities and the labor force transition probabilities as follows:

Theorem 1. *Under Assumptions 1 to 7, the misclassification probabilities in periods t and $t+1$, i.e., $\Pr\left(S_t | S_t^*, S_{t-1}, \boldsymbol{X}\right)$ and $\Pr\left(S_{t+1} | S_{t+1}^*, S_t, \boldsymbol{X}\right)$, as well as the labor force transition probabilities, i.e., $\Pr\left(S_{t+1}^* | S_t^*, S_{t-1}, \boldsymbol{X}\right)$, are uniquely identified from the observed joint distribution of five-period matched reported labor force statuses, i.e., $\Pr\left(S_{t+10}, S_{t+1}, S_t, S_{t-1}, S_{t-2} | \boldsymbol{X}\right)$, through using the eigenvalue-eigenvector decomposition twice.*

Proof. See Section B in the online Appendix. □

4 Data, Estimation and Simulations

4.1 Data

We use the public-use CPS data from January 1996 to December 2019. Due to the 4-8-4 rotating structure, we are able to match the CPS sample to form short longitudinal panels and obtain the joint distributions of the observed labor force statuses. We use the algorithm proposed by Madrian and Lefgren (2000) to match the CPS monthly files, as in Feng and Hu (2013) and Feng et al. (2022). Specifically, we first match the CPS sample based on household identifier, household replacement number and personal identifier, and then use information on sex, age, and race to "certify" the match in the first step.

The matched sample is not representative of the cross-sectional sample in month t due to sample attrition (Feng 2008). Therefore, we produce a matching weight based on the probability of being matched to correct for sample attrition. We first run a logit regression, where the dependent variable is whether an individual is matched or not and the independent variables are age, gender, race, education, region and labor force status in month t. Second, we predict the probabilities of being matched for all the observations in the matched sample. Finally, the matched sample is then weighted using the inverse of the predicted probabilities of being matched.

[14] We denote $M_{S_t^*, s_{t-1}, S_{t-2}|\mathbf{x}}$ as the matrix form of $\Pr\left(S_t^*, S_{t-1} = s_{t-1}, S_{t-2} | \mathbf{x}\right)$.

4.2 Specification for estimation with observed characteristics

It is important to control for observed heterogeneity, because people with different characteristics might have distinct misreporting behavior and labor force dynamics. Ideally, we would use identification results in Section 3 to obtain the closed-form estimation results of misclassification probabilities and transition probabilities for each demographic group non-parametrically. However, applying the estimation procedure to each subsample would lead to potential corner solutions, because we are dealing with the joint distributions of five-period matched data and some cells have extremely small sample sizes. For this reason, we instead utilize parametric Maximum Likelihood Estimation (MLE) with observed characteristics in the empirical analysis.

Based on Equations (1) and (6), we derive the following joint distribution with observed characteristics:

$$
\begin{aligned}
& \Pr\left(S_{t+10}, S_{t+1}, S_t, S_{t-1}, S_{t-2}, \mathbf{X}\right) \\
= & \sum_{S^*_{t+1}, S^*_t} \Pr\left(S_{t+10} | S^*_{t+1}, S_t, \mathbf{X}\right) \Pr\left(S_{t+1} | S^*_{t+1}, S_t, \mathbf{X}\right) \Pr\left(S^*_{t+1} | S^*_t, S_{t-1}, \mathbf{X}\right) \Pr\left(S_t | S^*_t, S_{t-1}, \mathbf{X}\right) \times \\
& \Pr\left(S^*_t | S_{t-1}, S_{t-2}, \mathbf{X}\right) \Pr\left(S_{t-1}, S_{t-2}, \mathbf{X}\right).
\end{aligned}
$$

We use a multinomial logit specification to estimate conditional probabilities with discrete choices involved, as in Poterba and Summers (1995). The observed covariates enter into this specification in a form of linear index. For example, the misclassification probabilities are parameterized as follows:

$$
\Pr\left(S_\tau = j | S^*_{\tau,} = k, S_{\tau-1} = s_{\tau-1}, \mathbf{X} = \mathbf{x}\right) = \frac{\exp\left(\alpha_{\tau,j,k} + s'_{\tau-1}\beta_{j,k} + \mathbf{x}'\gamma_{j,k}\right)}{\sum_{l=1}^{3} \exp\left(\alpha_{\tau,j,l} + s'_{\tau-1}\beta_{j,l} + \mathbf{x}'\gamma_{j,l}\right)}, \tag{7}
$$

for $j \in \{1, 2, 3\}$ and $\tau \in \{t, t+1\}$, where $\{\alpha_{\tau,j,k}, \beta_{j,k}, \gamma_{j,k}\}$ are unknown parameters with $\alpha_{\tau,3,k} = \beta_{3,k} = \gamma_{3,k} = 0$ for normalization. Notice that S^*_τ enters into the linear index non-parametrically, which is reflected in the second subscript in each parameter.[15] $S_{\tau-1}$ is a set of dummies for the reported labor force status, and the observed characteristics $\mathbf{X}$ include gender (male and female), race (white and nonwhite), while age groups (aged 16–24, 25–54, and 55 plus) are controlled for non-parametrically as they are highly collinear with the labor force statuses. Besides, we also include a set of dummies for five sub-periods according to NBER-based business cycles, including 1996Q1–2000Q4, 2001Q1–2001Q4, 2002Q1–2007Q3, 2007Q4–2009Q2, and 2009Q3–2019Q4.[16] In terms of the non-stationarity of misclassification probabilities, we only allow the constants (i.e., $\alpha_{\tau,j,k}$) to be different in periods t and $t+1$, but force all other parameters in $\Pr\left(S_{t+1} | S^*_{t+1}, S_t, \mathbf{X}\right)$ and $\Pr\left(S_t | S^*_t, S_{t-1}, \mathbf{X}\right)$ to be equal to

[15] There would be a multicollinearity problem if both S^*_τ and S_τ are included parametrically, as they are highly correlated.

[16] See https://www.nber.org/research/business-cycle-dating.

reduce estimation burden.[17] The other three conditional probabilities are parameterized in a similar fashion. In total, this gives us 666 unknown parameters to be estimated.

We then correct the monthly transition probabilities using the estimated misclassification probabilities. In order to increase sample size, we use three-month matched CPS data. Consider the following equation:

$$
\begin{aligned}
& \Pr\left(S_{t+1}, S_t, S_{t-1}, \mathbf{X}\right) \\
= & \sum_{S_{t+1}^*, S_t^*} \Pr\left(S_{t+1} | S_{t+1}^*, S_t, \mathbf{X}\right) \Pr\left(S_t | S_t^*, S_{t-1}, \mathbf{X}\right) \Pr\left(S_{t+1}^*, S_t^*, S_{t-1} | \mathbf{X}\right) \Pr\left(\mathbf{X}\right).
\end{aligned} \tag{8}
$$

Since the misclassification probabilities in periods t and $t+1$ have been estimated, we may obtain $\Pr\left(S_{t+1}^*, S_t^*, S_{t-1} | \mathbf{X}\right)$ through minimizing the Euclidean distance between the left-hand and right-hand sides of Equation (8). For the overall transition probabilities, we can simply integrate out the covariates as follows:

$$
\Pr\left(S_{t+1}^* | S_t^*\right) = \frac{\sum_{S_{t-1}, \mathbf{X}} \Pr\left(S_{t+1}^*, S_t^*, S_{t-1} | \mathbf{X}\right) \Pr\left(\mathbf{X}\right)}{\sum_{S_{t+1}^*, S_{t-1}, \mathbf{X}} \Pr\left(S_{t+1}^*, S_t^*, S_{t-1} | \mathbf{X}\right) \Pr\left(\mathbf{X}\right)}.
$$

4.3 Simulations

We next investigate the performance of our proposed estimation method using simulated data. We consider several cases with different true data generating processes (DGPs). For each case, we show three estimators. The first one is directly calculated from the reported data, which ignores misclassification errors. The second one is based on the restrictive method with strong assumptions imposed, i.e., the ICE assumption, the first-order Markov assumption, and the stationarity assumption. The third one is based on our proposed method. We summarize the main results here and leave all the details in Section D in the online Appendix.

In Case 1, we let the true DGP satisfy the assumptions proposed in this paper, and the simulation results for this case are shown in Table D1 in the online Appendix. The reported transition probabilities are all significantly biased, and the restrictive method produces even larger biases because it misspecifies the misclassification process and the underlying true labor force dynamics. On the contrary, our method substantially reduces biases, although the standard deviations are much larger. Overall, in terms of Mean Squared Error (MSE), our estimator performs much better than the restrictive one. In Case 2, we consider a more restrictive DGP that the latent labor force status follows the first-order Markov process and the misclassification process satisfies the local independence

[17] In Section 5.1, we also provide results that allow all the parameters to be different, which show our baseline results are robust.

assumption and the stationarity assumption. Table D2 shows that both our proposed method and the restrictive one perform well in correcting for biases in transition probabilities. As expected, in this case, the MSE of our estimator is in general slightly larger than the restrictive one. Taken together, the simulation results for both cases indicate that our proposed method can handle cases where the DGPs are more restrictive than those satisfying our maintained assumptions.

Furthermore, we use simulated data to examine if our proposed method performs well when the true DGP is more complicated. In the following two cases, we include more lags of the underlying true status in the conditional probabilities to account for the possibilities of unobserved heterogeneity or labor market persistence. In Case 3, we allow our Assumption 1 to deviate by letting the misclassification probabilities further depend on the true status in period $t-1$, i.e., $\Pr\left(S_t|S_t^*, \{S_\tau, S_\tau^*\}_{\tau\le t-1}, \mathbf{X}\right) = \Pr\left(S_t|S_t^*, S_{t-1}, S_{t-1}^*, \mathbf{X}\right)$. While in Case 4, we allow our Assumption 2 to deviate by letting the labor transition probabilities across periods t and $t+p$ further depend on the true status in period $t-1$, i.e., $\Pr\left(S_{t+p}^*|\{S_\tau, S_\tau^*\}_{\tau\le t}, \mathbf{X}\right) = \Pr\left(S_{t+p}^*|S_t^*, S_{t-1}, S_{t-1}^*, \mathbf{X}\right)$. In Sections D.2 and D.3 in the online Appendix, the simulation results show that our estimator is still robust to reasonable deviations from Assumptions 1 and 2 and significantly outperforms the reported one and the restrictive one.

5 Empirical Results

5.1 Misclassification probabilities

In order to incorporate heterogeneity among different demographic groups and sub-periods, we condition our estimates on gender, race, three age groups, and five sub-periods (see Table G1 in the online Appendix). The misclassification probabilities for each demographic subgroup are reported in Tables G2 and G3 in the online Appendix. We observe substantial differences across demographic groups.[18] In order to display the patterns clearly, we also report misclassification probabilities for the whole sample.[19]

Panels A and B of Table 1 show the misclassification probabilities in periods t and $t+1$, i.e., $\Pr\left(S_t|S_t^*, S_{t-1}\right)$ and $\Pr\left(S_{t+1}|S_{t+1}^*, S_t\right)$, respectively. On the one hand, we do find persistence in the misclassification process in both periods, similar to earlier studies such as Keane and Sauer (2010) and Feng et al. (2022). In addition, the ordering patterns of the columns of the misclassification probabilities matrices are exactly the same as described in Assumption 5. For example, in period t,

[18] Formal tests are reported in Tables G4, G5, G6, and G7 in the online Appendix.

[19] The estimates for the whole sample can be calculated as follows:

$$\Pr\left(S_t|S_t^*, S_{t-1}\right) = \frac{\Pr\left(S_t, S_t^*, S_{t-1}\right)}{\Pr\left(S_t^*, S_{t-1}\right)} = \frac{\sum_{S_{t-2},\mathbf{X}} \Pr\left(S_t|S_t^*, S_{t-1}, \mathbf{X}\right) \Pr\left(S_t^*|S_{t-1}, S_{t-2}, \mathbf{X}\right) \Pr\left(S_{t-1}, S_{t-2}, \mathbf{X}\right)}{\sum_{S_t, S_{t-2},\mathbf{X}} \Pr\left(S_t|S_t^*, S_{t-1}, \mathbf{X}\right) \Pr\left(S_t^*|S_{t-1}, S_{t-2}, \mathbf{X}\right) \Pr\left(S_{t-1}, S_{t-2}, \mathbf{X}\right)}.$$

given that the previously-reported status is E, the probability of being classified as E is 98.2% for truly employed people, higher than those who are truly unemployed (77.5%) and truly not-in-the-labor-force (65.2%). Furthermore, conditional on that the previously-reported status is E, the probability of being misclassified as N is 7.3% for truly unemployed people, and the probability of being misclassified as U is 3.6% for those who are truly not-in-the-labor-force. Both of the two numbers are the smallest entries in their respective columns. These results show that not only the current true status, but also the previously-reported status affect the current reported labor force status. Therefore, one cannot describe the misclassification process simply by using a 3-by-3 matrix and assuming that people are more likely to tell the truth (i.e., the diagonal elements are the largest in each column of that 3-by-3 matrix), especially in the context of studying labor force dynamics, as previous studies did (e.g., Biemer and Bushery 2000).

On the other hand, we also find that the misclassification probabilities in periods t and $t+1$ are significantly different in magnitudes, as shown in Panel C of Table 1. This suggests that the stationarity assumption imposed on the misclassification process that has been widely used in the existing studies might be too strong (e.g., Biemer and Bushery 2000; Shibata 2022). More importantly, the diagonal elements in $\Pr\left(S_{t+1}|S_{t+1}^{*}, S_t\right)$ are significantly larger than those in $\Pr\left(S_t|S_t^{*}, S_{t-1}\right)$, which also holds for all demographic groups (see Table G8 in the online Appendix). This implies that the misclassification errors in the fourth interview are smaller than those in the third one, which is consistent with a learning mechanism at work.[20]

Finally, we perform two robustness checks and find that the estimated misclassification probabilities are qualitatively unchanged, and all the patterns we discussed before still hold. In the first exercise, we allow the parameterization to be more flexible in the multinomial logit specification reported in Equation (7), so that all parameters are conditional on time, i.e., using $\beta_{\tau,j,k}$ and $\gamma_{\tau,j,k}$ instead of $\beta_{j,k}$ and $\gamma_{j,k}$. The results are reported in Table G9 in the online Appendix. In the second exercise, we add a fourth labor force status, "missing", in addition to E, U and N to the fifth period of matched data as an alternative way to take care of the sample attrition problem.[21] The estimated results are reported in Table G10 in the online Appendix.

[20] Bollinger and David (2005) study response errors in Food Stamp program participation. They identify persistent response errors but do not find evidence for a learning mechanism, although their study contexts are quite different from this paper.

[21] The identification argument for this case is the same as that described in Section 3.2. The reason why we focus on the fifth period (rotation group five) of the matched data is that the sample attrition issue becomes more serious after the eight-month drop-out periods. When matching the first four monthly data (rotation groups one to four), only about 10% sample are lost due to attrition. But when the data from rotation group five are also matched together with the first four periods, the sample attrition rate increases significantly to about 30%.

5.2 Transition probabilities

We then present our estimates of transition probabilities.[22] Figure 1 shows the reported and the corrected transition probabilities. For the diagonal panels, which show the probabilities of staying on the same labor force statuses, we see that our corrected transition probabilities are all lower than the reported ones. The opposite is true for the off-diagonal panels. Therefore, our method shows that compared to what the raw data suggests, there is actually less persistence, or more mobility, in individuals' labor force statuses across months.

We also average these monthly transition probabilities and summarize them in Table 2. In Panel A, we show the transition probabilities conditional on reported status in the previous period. There is indeed spurious transitions in the raw data, as the reported and the corrected transition probabilities are different in all cases, with the differences statistically significant. The magnitudes of the differences differ substantially across the previously-reported statuses. For example, when the reported status is U in period $t-1$, the reported E-to-U transition probability from period t to $t+1$ is 11.7%, but 51.7% according to our corrected estimates. This adjustment is much larger than when the reported status is E or N in period $t-1$. This also confirms the importance of controlling for historical labor force status when estimating transition probabilities. In Panel B of Table 2, we integrate out the previously-reported status and report the overall transition probabilities across two consecutive months. We find that the reported E-to-E, U-to-U and N-to-N transitions are overestimated, consistent with what we observe in Figure 1. Accordingly, our corrected estimates show higher labor mobility, as the other off-diagonal probabilities increase after correction, except for the U-to-N transition. Overall, these patterns also hold for different demographic groups, which are shown in Table G11 in the online Appendix.

In Table 3, we formally test the first-order Markov assumption imposed on the underlying true labor force dynamics. We find that the transition probabilities between t and $t+1$ differ significantly across the reported statuses in period $t-1$, suggesting that we can reject the first-order Markov assumption.[23] More importantly, we find that $\Pr\left(S^*_{t+1} = i | S^*_t = i, S_{t-1} = i, \mathbf{X}\right) > \Pr\left(S^*_{t+1} = i | S^*_t = i, S_{t-1} = j, \mathbf{X}\right)$ for $i \neq j$, lending support to the existence of higher-order state dependence, which is one of the non-Markovian characteristics of the true labor force dynamics. For example, unemployed people with a previously-reported status of U have a higher tendency of staying unemployed than those with a

[22] In Figures G1, G2 and G3 in the online Appendix, we plot the reported and the corrected unemployment rate, labor force participation rate and employment-to-population ratio, respectively. We also compare our estimates with those based on Feng and Hu (2013) and Feng et al. (2022). For unemployment rate, all the three corrected lines are higher than the reported one, confirming that the reported (official) unemployment rate is underestimated. The three corrected lines are somewhat different in levels, partly because they are calculated using different identification strategies and data information. Nonetheless, all the three corrected lines are evolving quite similarly with the reported one.

[23] Similarly, based on the estimates of $\Pr\left(S^*_{t+1} | S^*_t, S_{t-1}, \mathbf{X}\right)$ in Table G12 in the online Appendix, we perform the same tests for all demographic groups in Table G13 and reject the first-order Markov assumption consistently.

previously-reported status of E or N. In the meanwhile, people who previously reported U have the lowest U-to-E transition probability and the highest E-to-U transition probability. These results are also related to the literature on duration dependence, such as Krueger et al. (2014), which find that people who have experienced longer duration of unemployment are less likely to find jobs, and when they find jobs, they are still more likely to lose jobs and return to unemployment.[24]

In Table 4, we compare our estimates with those in the existing studies that also correct for misclassification errors in gross labor flows using the CPS data. Among them, the "Reported" estimates are calculated directly from the raw data (e.g., Shimer 2012), the "Corrected" are estimated using our proposed framework, while the "Corrected-FH" is based on the framework in Feng and Hu (2013) with small modifications.[25] The "Corrected-AZ", "Corrected-PS" and "Corrected-S" are based on the estimates of misclassification probabilities from Abowd and Zellner (1985), Poterba and Summers (1986), and Shibata (2022), respectively. It is striking that our method differs from all the other methods in terms of the direction of adjustments in the reported transition probabilities. For the diagonal elements of the transition matrix, our estimates are lower than the reported ones, while the other estimates are all higher. Overall, our method implies that the raw data understates the actual labor market mobility, while the other methods suggest otherwise. This is because that none of the existing methods have considered persistence in measurement errors and underlying labor force dynamics as we do. Under our framework, these restrictive methods would inappropriately attribute all serial correlations in the observed data to the true underlying labor force dynamics, thus overestimate persistence and underestimate mobility of individual labor force statuses.[26] The intuition here is similar to the case when classical measurement errors attenuate estimated coefficients in a linear regression framework.

5.3 Discussions

In this paper, we have developed a generalized model of the misclassification process and the latent true labor force dynamics, extending the existing framework along different dimensions. First, for the misclassification process, we relax the strong ICE assumption and consider potential correlation in measurement errors across time, allowing the misclassification errors to not only depend on current true status, but also on previously-reported status. Second, we also allow the misclassification probabilities to be different for two consecutive months, dropping the stationarity assumption. Lastly,

[24]Our results are not completely comparable to this strand of literature, as we analyze labor market persistence across a three-month period after correcting for misclassification errors, while others mainly use the observed data in a longer-term view.

[25]When correcting labor transition probabilities using the framework of Feng and Hu (2013), we allow for non-stationarity of the misclassification probabilities. See detailed modifications in Section F in the online Appendix.

[26]van de Pol and Langeheine (1997) and Bassi and Trivellato (2009) report similar findings as ours using data from European countries.

for the underlying true dynamics of labor force status, we extend the first-order Markov assumption and allow the latent transition probabilities to also depend on the reported status in the previous period. In all cases, we have empirically rejected the assumptions of ICE, stationarity and first-order Markov. However, it is still not clear which extensions are more important in driving our new estimates quantitatively. We try to answer that question in this subsection.

We start with the restrictive framework with all three assumptions imposed. Under the ICE assumption for the misclassification process and the first-order Markov assumption for the underlying true labor force dynamics, we can derive the following equation:

$$\begin{aligned} & \Pr\left(S_{t+10}, S_{t+1}, S_t, S_{t-1}, S_{t-2}, \mathbf{X}\right) \\ = & \sum_{S_{t+1}^*, S_t^*} \Pr\left(S_{t+10}|S_{t+1}^*, \mathbf{X}\right) \Pr\left(S_{t+1}|S_{t+1}^*, \mathbf{X}\right) \Pr\left(S_{t+1}^*|S_t^*, \mathbf{X}\right) \Pr\left(S_t|S_t^*, \mathbf{X}\right) \times \\ & \Pr\left(S_t^*|S_{t-1}, S_{t-2}, \mathbf{X}\right) \Pr\left(S_{t-1}, S_{t-2}, \mathbf{X}\right) \end{aligned}$$

With the stationarity assumption imposed as well, i.e., $\Pr\left(S_{t+1}|S_{t+1}^*\mathbf{X}\right) = \Pr\left(S_t|S_t^*, \mathbf{X}\right)$, we can estimate the misclassification probabilities and the transition probabilities using the same procedures as we outlined before in the paper. The estimated transition probabilities are reported in Table 5 as "Restrictive".

We then consider relaxing the three assumptions. In each case, we only keep one of the three assumptions and extend the other two to the specification that we use in our proposed model. For example, "ICE" means that we only keep the ICE assumption for the misclassification process, and "Markov" stands for the case when we only impose the first-order Markov assumption for the underlying labor force dynamics. Similarly, "Stationarity" means we only force the two misclassification probabilities to be the same. Finally, "Baseline" refers to the results from this paper, in which all three assumptions are relaxed.

The results are reported in Table 5. First, note that as we have explained in the last subsection, our "Baseline" results correct the transition probabilities to the opposite direction of what the "Restrictive" ones do. The "Restrictive" transition probabilities are very similar to what was derived using previous models in the literature, as reported in Table 4. This shows that the previous estimates are indeed driven by the key assumptions that we highlight. Second, we see that the stationarity assumption imposed on the misclassification probabilities seems to have played only a minor role, as the "Stationarity" results are quite similar to our "Baseline" results. On the other hand, the other two assumptions are much more important quantitatively, as we see that the "ICE" and the "Markov" results are all very close to the "Restrictive", and quite different from the "Baseline". Overall, we can summarize from this table that persistence in both the misclassification process and the underlying labor force dynamics are important and need to be considered when analyzing gross labor flows.

Next, we consider a scenario where Assumptions 1 and 2 are simultaneously relaxed as follows:

$$\begin{aligned}
\Pr\left(S_t|S_t^*, \{S_\tau^*, S_\tau\}_{\tau\leq t-1}, \mathbf{X}\right) &= \Pr\left(S_t|S_t^*, S_{t-1}, S_{t-2}, \mathbf{X}\right), \\
\Pr\left(S_{t+p}^*| \{S_\tau, S_\tau^*\}_{\tau\leq t}, \mathbf{X}\right) &= \Pr\left(S_{t+p}^*|S_t^*, S_{t-1}, S_{t-2}, \mathbf{X}\right).
\end{aligned}$$

The relaxation of Assumption 1 is along the lines of Bollinger and David (2005) who study the higher-order persistence in the misclassification process. While Assumption 2 is relaxed in a manner consistent with Kudlyak and Lange (2018), which consider the impact of historical labor market experiences on labor force transition. In the last row of Table 5, we report the corrected transition probabilities based on this setting in "More lags", showing that they differ slightly from the baseline results but have larger standard errors. The corrected estimates still show more fluidity in labor force transition than the reported ones, confirming the robustness of our main baseline results. The misclassification probabilities in this setting are also quite similar to the baseline results, as shown in Tables C1 and C2 in the online Appendix. It is worth noting that given the 4-8-4 rotating structure of the CPS, this is the most general setting under which we can still show identification with the conditional independence assumption.[27] Adding more lags will lose the identification arguments. However, the costs are that we need to impose the stationarity assumption on the misclassification process, and cannot control for as many demographic characteristics as in the baseline setting.

6 Conclusion

Understanding gross labor flows is critical for macroeconomic and labor market policy-making. Although researchers have realized that misclassification errors in labor force status can bias estimates of gross labor flows, these errors have not been properly corrected for. The most recent studies typically make very strong assumptions on the measurement error structure and the underlying labor force dynamics. Specifically, measurement errors are assumed to be only dependent on current true status (the ICE assumption), and the same across two consecutive months (the stationarity assumption), and the underlying labor force dynamics are assumed to follow a first-order Markov process. These assumptions imply that the observed data would generate spurious transitions among different labor force statuses, and exaggerate the "true" mobility in labor force statuses. Therefore, the corrected estimates based on restrictive methods exhibit lower labor mobility than the observed ones.

In this paper, we develop a generalized model of the misclassification process and the underlying

[27] The identification arguments of this setting are presented in Section C in the online Appendix.

true labor force dynamics by relaxing the restrictive assumptions used in the literature. Importantly, we allow for correlation across time in both the misclassification process and the underlying true labor force dynamics. Regarding the misclassification process, we follow Feng et al. (2022) to allow the current reported status to depend not only on the current true status, but also on the previously-reported status. We further allow the misclassification probabilities to be different in two consecutive months. With respect to the underlying labor force dynamics, we allow the transition probabilities to be also affected by the previously-reported status. Following Hu (2008), we establish a closed-form global identification results using an eigenvalue-eigenvector decomposition under certain reasonable and testable assumptions. We also incorporate heterogeneity among different demographic groups using a flexible multinomial logit specification.

Using the CPS data, we estimate the misclassification probabilities and empirically reject the ICE assumption. We also find that the misclassification probabilities in two consecutive months are different, thus rejecting the stationarity assumption. Also, we find that the underlying labor force dynamics do not follow a simple first-order Markov process. Rather, the transition probabilities from period t to period $t+1$ depend on the reported status in period $t-1$, suggesting that historical labor market experience might influence labor force transitions. We further find empirical evidence on higher-order state dependence. Contrary to existing studies, our results suggest that the observed data have understated the true mobility in labor force statuses after we account for persistence in both the measurement errors and the true underlying dynamics.

Although our modeling framework is specifically tailored to the structure of the CPS, we believe that it can be applied more generally to other panel datasets with suitable modifications. Also, the findings of persistent misclassification errors and higher-order state dependence are unlikely to be only applicable in the U.S. context. Indeed, van de Pol and Langeheine (1997) and Bassi and Trivellato (2009) have reached qualitatively similar conclusions using different methodologies with European datasets. Our results show that it is important to properly account for persistence in both the misclassification errors and the latent true labor force dynamics when studying labor force flows.

References

Abowd, John M., and Arnold Zellner. 1985. "Estimating gross labor-force flows." *Journal of Business & Economic Statistics* 3 (3): 254–283.

Abraham, Katharine G., John Haltiwanger, Kristin Sandusky, and James R. Spletzer. 2013. "Exploring differences in employment between household and establishment data." *Journal of Labor Economics* 31 (S1): S129–S172.

Ahn, Hie Joo, and James D. Hamilton. 2020. "Heterogeneity and unemployment dynamics." *Journal of Business & Economic Statistics* 38 (3): 554–569.

———. 2022. "Measuring labor-force participation and the incidence and duration of unemployment." *Review of Economic Dynamics* 44:1–32.

Ahn, Hie Joo, Bart Hobijn, and Ayşegül Şahin. 2023. *The dual U.S. labor market uncovered.* NBER Working Paper 31241. National Bureau of Economic Research.

Bassi, Francesca, and Ugo Trivellato. 2009. "A latent class approach for estimating gross flows in the presence of correlated classification errors." In *Methodology of Longitudinal Surveys,* edited by Lynn Peter, 367–380. Chichester: John Wiley & Sons.

Biemer, Paul P., and John M. Bushery. 2000. "On the validity of Markov latent class analysis for estimating classification error in labor force data." *Survey Methodology* 26 (2): 139–152.

Biemer, Paul P., and Gösta Forsman. 1992. "On the quality of reinterview data with application to the Current Population Survey." *Journal of the American Statistical Association* 87 (420): 915–923.

Bollinger, Christopher R., and Martin H. David. 2005. "I didn't tell, and I won't tell: Dynamic response error in the SIPP." *Journal of Applied Econometrics* 20 (4): 563–569.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement error in survey data." In *Handbook of Econometrics Volume 5,* edited by James J. Heckman and Edward Leamer, 3705–3843. Amsterdam: Elsevier.

Burdett, Kenneth, Nicholas M. Kiefer, Dale T. Mortensen, and George R. Neumann. 1984. "Earnings, unemployment, and the allocation of time over time." *The Review of Economic Studies* 51 (4): 559–578.

Burdett, Kenneth, Nicholas M. Kiefer, and Sunil Sharma. 1985. "Layoffs and duration dependence in a model of turnover." *Journal of Econometrics* 28 (1): 51–69.

Chua, Tin Chiu, and Wayne A. Fuller. 1987. "A model for multinomial response error applied to labor flows." *Journal of the American Statistical Association* 82 (397): 46–51.

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

Eriksson, Stefan, and Dan-Olof Rooth. 2014. "Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment." *American Economic Review* 104 (3): 1014–39.

Farber, Henry S., and Robert G. Valletta. 2015. "Do extended unemployment benefits lengthen unemployment spells? Evidence from recent cycles in the US labor market." *Journal of Human Resources* 50 (4): 873–909.

Feng, Shuaizhang. 2008. "Longitudinal matching of recent Current Population Surveys: Methods, non-matches and mismatches." *Journal of Economic and Social Measurement* 33 (4): 241–252.

Feng, Shuaizhang, and Yingyao Hu. 2013. "Misclassification errors and the underestimation of the US unemployment rate." *American Economic Review* 103 (2): 1054–1070.

Feng, Shuaizhang, Yingyao Hu, and Jiandong Sun. 2022. "Rotation group bias and the persistence of misclassification errors in the Current Population Surveys." *Econometric Reviews* 41 (9): 1077–1094.

Flinn, Christopher, and James J. Heckman. 1982. "New methods for analyzing structural models of labor force dynamics." *Journal of Econometrics* 18 (1): 115–168.

Gregory, Victoria, Guido Menzio, and David G. Wiczer. 2021. *The Alpha Beta Gamma of the labor market.* NBER Working Paper 28663. National Bureau of Economic Research.

Hall, Robert E., and Marianna Kudlyak. 2019. *Job-finding and job-losing: A comprehensive model of heterogeneous individual labor-market dynamics.* NBER Working Paper 25625. National Bureau of Economic Research.

Halpern-Manners, Andrew, and John Robert Warren. 2012. "Panel conditioning in the Current Population Survey: Implications for labor force statistics." *Demography* 49:1499–1519.

Heckman, James J. 1981a. "Heterogeneity and state dependence." In *Studies in Labor Markets,* edited by Sherwin Rosen, 91–140. Chicago: University of Chicago Press.

———. 1981b. "Statistical models for discrete panel data." In *Structural Analysis of Discrete Data with Econometric Applications,* edited by Charles F. Manski and Daniel McFadden, 114–178. Boston: MIT Press.

Heckman, James J., and George J. Borjas. 1980. "Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence." *Economica* 47 (187): 247–283.

Heckman, James J., and Robert J. Willis. 1977. "A beta-logistic model for the analysis of sequential labor force participation by married women." *Journal of Political Economy* 85 (1): 27–58.

Hu, Yingyao. 2008. "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution." *Journal of Econometrics* 144 (1): 27–61.

———. 2017. "The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics." *Journal of Econometrics* 200 (2): 154–168.

Hyslop, Dean R. 1999. "State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women." *Econometrica* 67 (6): 1255–1294.

Keane, Michael P., and Robert M. Sauer. 2010. "A computationally practical simulation estimation algorithm for dynamic panel data models with unobserved endogenous state variables." *International Economic Review* 51 (4): 925–958.

Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo. 2013. "Duration dependence and labor market conditions: Evidence from a field experiment." *Quarterly Journal of Economics* 128 (3): 1123–1167.

Kroft, Kory, Fabian Lange, Matthew J. Notowidigdo, and Lawrence F. Katz. 2016. "Long-term unemployment and the Great Recession: The role of composition, duration dependence, and non-participation." *Journal of Labor Economics* 34 (S1): S7–S54.

Krueger, Alan B., Judd Cramer, and David Cho. 2014. "Are the long-term unemployed on the margins of the labor market?" *Brookings Papers on Economic Activity,* no. 1, 229–299.

Kudlyak, Marianna, and Fabian Lange. 2018. *Measuring heterogeneity in job finding rates among the non-employed using labor force status histories.* Working Paper 2017-20. Federal Reserve Bank of San Francisco.

Madrian, Brigitte C., and Lars John Lefgren. 2000. "An approach to longitudinally matching Current Population Survey (CPS) respondents." *Journal of Economic and Social Measurement* 26 (1): 31–62.

Magnac, Thierry. 2000. "Subsidised training and youth employment: Distinguishing unobserved heterogeneity from state dependence in labour market histories." *Economic Journal* 110 (466): 805–837.

Magnac, Thierry, and Michael Visser. 1999. "Transition models with measurement errors." *Review of Economics and Statistics* 81 (3): 466–474.

Meyer, Bruce D. 1988. "Classification-error models and labor-market dynamics." *Journal of Business & Economic Statistics* 6 (3): 385–390.

Poterba, James M., and Lawrence H. Summers. 1986. "Reporting errors and labor market dynamics." *Econometrica* 54 (6): 1319–1338.

———. 1995. "Unemployment benefits and labor market transitions: A multinomial logit model with errors in classification." *Review of Economics and Statistics* 77 (2): 207–216.

Prowse, Victoria. 2012. "Modeling employment dynamics with state dependence and unobserved heterogeneity." *Journal of Business & Economic Statistics* 30 (3): 411–431.

Rothstein, Jesse. 2011. "Unemployment insurance and job search in the Great Recession." *Brookings Papers on Economic Activity,* no. 2, 143–196.

Shibata, Ippei. 2019. *Labor market dynamics: A hidden Markov approach.* Working Paper 2019/282. International Monetary Fund.

———. 2022. "Reassessing classification errors in the analysis of labor market dynamics." *Labour Economics* 78:102252.

Shimer, Robert. 2012. "Reassessing the ins and outs of unemployment." *Review of Economic Dynamics* 15:127–148.

Sinclair, Michael D., and Joseph L. Gastwirth. 1996. "On procedures for evaluating the effectiveness of reinterview survey methods: Application to labor force data." *Journal of the American Statistical Association* 91 (435): 961–969.

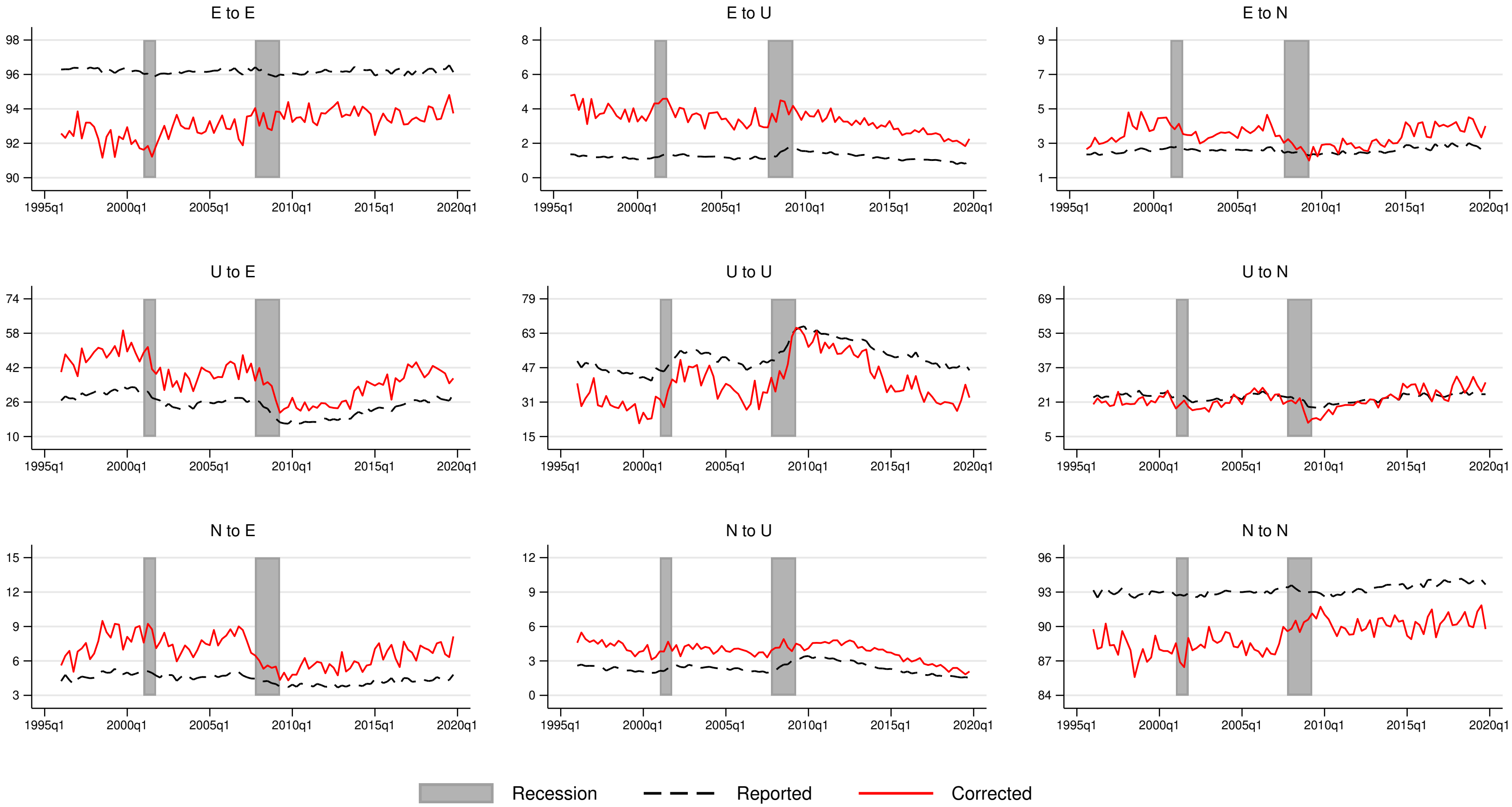
Singh, Avinash C., and J. N. K. Rao. 1995. "On the adjustment of gross flow estimates for classification error with application to data from the Canadian Labour Force Survey." *Journal of the American Statistical Association* 90 (430): 478–488.

Toikka, Richard S. 1976. "A Markovian model of labor market decisions by workers." *American Economic Review* 66 (5): 821–834.

van de Pol, Frank, and Rolf Langeheine. 1997. "Separating change and measurement error in panel surveys with an application to labor market data." In *Survey Measurement and Process Quality,* edited by Lars Lyberg, Paul Biemer, Martin Collins, Edith de Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin, 671–688. New York: John Wiley & Sons.

van den Berg, Gerard J., and Jan C. van Ours. 1996. "Unemployment dynamics and duration dependence." *Journal of Labor Economics* 14 (1): 100–125.

Figure 1: Reported and corrected transition probabilities, 1996-2019



Note: Reported and corrected transition probabilities, quarterly average of monthly data. The monthly gross flow are seasonally adjusted using a ratio to moving average.

Table 1: Misclassification probabilities

	$k=E$			$k=U$			$k=N$		
	$j=E$	$j=U$	$j=N$	$j=E$	$j=U$	$j=N$	$j=E$	$j=U$	$j=N$
Panel A: $\Pr\left(S_t=i \mid S_t^*=j, S_{t-1}=k\right)$									
$i=E$	98.2	77.5	65.2	48.4	9.8	10.9	35.5	8.3	1.4
	(0.02)	(0.82)	(0.31)	(0.73)	(0.28)	(0.32)	(0.33)	(0.31)	(0.02)
$i=U$	0.6	15.3	3.6	40.6	74.8	41.8	5.4	32.5	0.9
	(0.01)	(0.60)	(0.12)	(0.67)	(0.34)	(0.46)	(0.21)	(0.59)	(0.02)
$i=N$	1.1	7.3	31.2	11.0	15.4	47.4	59.1	59.2	97.7
	(0.01)	(0.30)	(0.32)	(0.30)	(0.27)	(0.50)	(0.29)	(0.62)	(0.03)
Panel B: $\Pr\left(S_{t+1}=i \mid S_{t+1}^*=j, S_t=k\right)$									
$i=E$	98.8	73.3	55.7	58.6	8.0	9.4	45.9	7.8	0.9
	(0.02)	(0.45)	(0.28)	(0.46)	(0.41)	(0.31)	(0.26)	(0.35)	(0.01)
$i=U$	0.4	19.7	2.4	32.5	79.8	26.4	4.7	39.2	0.5
	(0.01)	(0.34)	(0.09)	(0.42)	(0.52)	(0.45)	(0.14)	(0.43)	(0.01)
$i=N$	0.7	7.0	41.9	8.9	12.2	64.2	49.4	53.0	98.6
	(0.01)	(0.23)	(0.29)	(0.27)	(0.27)	(0.48)	(0.24)	(0.48)	(0.02)
Panel C: Testing the stationarity assumption, $\Pr\left(S_{t+1} \mid S_{t+1}^*, S_t\right) - \Pr\left(S_t \mid S_t^*, S_{t-1}\right)$									
$i=E$	0.6***	-4.2***	-9.5***	10.2***	-1.8***	-1.4***	10.4***	-0.5	-0.5***
	(0.02)	(0.74)	(0.32)	(0.50)	(0.40)	(0.17)	(0.32)	(0.44)	(0.02)
$i=U$	-0.2***	4.4***	-1.2***	-8.1***	5.0***	-15.4***	-0.7***	6.7***	-0.5***
	(0.01)	(0.53)	(0.08)	(0.56)	(0.54)	(0.48)	(0.15)	(0.61)	(0.02)
$i=N$	-0.4***	-0.2	10.7***	-2.1***	-3.2***	16.8***	-9.7***	-6.2***	1.0***
	(0.01)	(0.26)	(0.31)	(0.14)	(0.29)	(0.44)	(0.31)	(0.56)	(0.03)

Note: In parentheses are bootstrapped standard errors based on 500 repetitions. ***, **, * signify significance at the 1%, 5%, 10% level, respectively.

Table 2: Transition probabilities, averaged over 1996–2019

	Panel A: $\Pr(\mathscr{S}_{t+1} = i \mid \mathscr{S}_t = j, S_{t-1} = k)$								
	$(E\|E,k)$	$(U\|E,k)$	$(N\|E,k)$	$(E\|U,k)$	$(U\|U,k)$	$(N\|U,k)$	$(E\|N,k)$	$(U\|N,k)$	$(N\|N,k)$
(1) $k = E$									
Reported	97.1	1.0	1.9	44.4	40.4	15.2	34.4	6.2	59.4
	(0.01)	(0.00)	(0.00)	(0.14)	(0.14)	(0.11)	(0.09)	(0.05)	(0.09)
Corrected	97.2	1.7	1.1	56.2	29.7	14.1	55.9	13.0	31.1
	(0.06)	(0.06)	(0.03)	(1.08)	(0.84)	(0.43)	(0.64)	(0.34)	(0.52)
Difference	0.1*	0.7***	-0.8***	11.8***	-10.7***	-1.2***	21.5***	6.8***	-28.3***
	(0.06)	(0.06)	(0.02)	(1.05)	(0.81)	(0.43)	(0.61)	(0.33)	(0.49)
(2) $k = U$									
Reported	80.4	11.7	7.9	19.5	62.2	18.3	12.6	25.7	61.7
	(0.12)	(0.09)	(0.08)	(0.08)	(0.10)	(0.08)	(0.09)	(0.12)	(0.13)
Corrected	35.7	51.7	12.7	21.0	63.7	15.3	18.1	54.6	27.3
	(0.78)	(0.79)	(0.27)	(0.62)	(0.86)	(0.39)	(0.43)	(0.73)	(0.55)
Difference	-44.8***	40.0***	4.7***	1.5***	1.5*	-3.0***	5.5***	28.9***	-34.4***
	(0.76)	(0.78)	(0.26)	(0.60)	(0.84)	(0.37)	(0.40)	(0.70)	(0.53)
(3) $k = N$									
Reported	70.4	3.6	26.0	18.3	40.4	41.3	2.5	1.5	96.0
	(0.10)	(0.04)	(0.10)	(0.11)	(0.14)	(0.14)	(0.01)	(0.01)	(0.01)
Corrected	33.2	16.0	50.8	22.3	33.8	43.8	1.4	1.2	97.4
	(0.56)	(0.38)	(0.69)	(0.44)	(0.67)	(0.77)	(0.03)	(0.04)	(0.04)
Difference	-37.2***	12.4***	24.8***	4.0***	-6.5***	2.5***	-1.1***	-0.4***	1.5***
	(0.54)	(0.37)	(0.66)	(0.42)	(0.64)	(0.73)	(0.03)	(0.04)	(0.04)
	Panel B: $\Pr(\mathscr{S}_{t+1} = i \mid \mathscr{S}_t = j)$								
	$(E\|E)$	$(U\|E)$	$(N\|E)$	$(E\|U)$	$(U\|U)$	$(N\|U)$	$(E\|N)$	$(U\|N)$	$(N\|N)$
Reported	96.2	1.2	2.6	24.8	52.1	23.1	4.4	2.4	93.2
	(0.01)	(0.00)	(0.01)	(0.06)	(0.07)	(0.06)	(0.01)	(0.01)	(0.01)
Corrected	93.0	3.4	3.6	37.5	40.2	22.3	7.0	3.9	89.1
	(0.01)	(0.00)	(0.01)	(0.06)	(0.07)	(0.06)	(0.01)	(0.01)	(0.01)
Difference	-3.2***	2.2***	1.0***	12.7***	-11.8***	-0.9*	2.6***	1.5***	-4.0***
	(0.10)	(0.08)	(0.06)	(1.01)	(0.81)	(0.47)	(0.11)	(0.08)	(0.13)

Note: In parentheses are bootstrapped standard errors based on 500 repetitions. ***, **, * signify significance at the 1%, 5%, 10% level, respectively.

Table 3: Testing the first-order Markov assumption

	$\Delta_{i\|j}^{p-q}=\Pr\left(S_{t+1}^{*}=i\|S_{t}^{*}=j,S_{t-1}=p\right)-\Pr\left(S_{t+1}^{*}=i\|S_{t}^{*}=j,S_{t-1}=q\right)$								
	$\Delta_{E\|E}^{p-q}$	$\Delta_{U\|E}^{p-q}$	$\Delta_{N\|E}^{p-q}$	$\Delta_{E\|U}^{p-q}$	$\Delta_{U\|U}^{p-q}$	$\Delta_{N\|U}^{p-q}$	$\Delta_{E\|N}^{p-q}$	$\Delta_{U\|N}^{p-q}$	$\Delta_{N\|N}^{p-q}$
$p=E,\ q=U$	61.5***	-50.0***	-11.5***	35.2***	-34.0***	-1.2*	37.9***	-41.6***	3.8***
	(0.74)	(0.75)	(0.27)	(0.83)	(0.83)	(0.62)	(0.75)	(0.73)	(0.69)
$p=E,\ q=N$	64.0***	-14.3***	-49.6***	33.9***	-4.1***	-29.7***	54.6***	11.8***	-66.4***
	(0.55)	(0.36)	(0.68)	(1.03)	(0.94)	(0.85)	(0.63)	(0.34)	(0.51)
$p=U,\ q=N$	2.5***	35.6***	-38.1***	-1.3**	29.9***	-28.6***	16.7***	53.5***	-70.2***
	(0.84)	(0.74)	(0.67)	(0.61)	(0.86)	(0.73)	(0.42)	(0.70)	(0.54)

Note: In parentheses are bootstrapped standard errors based on 500 repetitions. ***, **, * signify significance at the 1%, 5%, 10% level, respectively.

Table 4: Comparing transition probabilities with existing studies

	$\Pr(\mathscr{S}_{t+1}=i \mid \mathscr{S}_t=j)$								
	$(E\|E)$	$(U\|E)$	$(N\|E)$	$(E\|U)$	$(U\|U)$	$(N\|U)$	$(E\|N)$	$(U\|N)$	$(N\|N)$
Reported	96.2	1.2	2.6	24.8	52.1	23.1	4.4	2.4	93.2
	(0.01)	(0.00)	(0.01)	(0.06)	(0.07)	(0.06)	(0.01)	(0.01)	(0.01)
Corrected	93.0	3.4	3.6	37.5	40.2	22.3	7.0	3.9	89.1
	(0.10)	(0.08)	(0.06)	(1.02)	(0.82)	(0.47)	(0.12)	(0.08)	(0.13)
Corrected-FH	98.5	0.6	0.9	11.8	80.7	7.5	1.1	0.7	98.3
	(0.01)	(0.00)	(0.00)	(0.08)	(0.11)	(0.08)	(0.01)	(0.01)	(0.01)
Corrected-AZ	97.7	1.1	1.3	21.7	62.7	15.6	2.1	1.7	96.1
	(0.01)	(0.00)	(0.01)	(0.07)	(0.09)	(0.07)	(0.01)	(0.01)	(0.01)
Corrected-PS	99.2	0.6	0.3	16.1	72.9	11.0	1.3	1.1	97.6
	(0.01)	(0.00)	(0.00)	(0.07)	(0.11)	(0.09)	(0.01)	(0.01)	(0.01)
Corrected-S	98.4	0.7	0.9	14.8	80.4	4.8	1.2	0.5	98.3
	(0.01)	(0.00)	(0.01)	(0.08)	(0.11)	(0.07)	(0.01)	(0.01)	(0.01)

Note: Average transition probabilities over 1996–2019. "Corrected-FH", "Corrected-AZ", "Corrected-PS", and "Corrected-S" are based on Feng and Hu (2013), Abowd and Zellner (1985), Poterba and Summers (1986), and Shibata (2022), respectively. In parentheses are bootstrapped standard errors based on 500 repetitions.

Table 5: Comparing transition probabilities with alternative restrictions

	$\Pr(\mathscr{S}_{t+1} = i \mid \mathscr{S}_t = j)$								
	$(E\|E)$	$(U\|E)$	$(N\|E)$	$(E\|U)$	$(U\|U)$	$(N\|U)$	$(E\|N)$	$(U\|N)$	$(N\|N)$
Reported	96.2	1.2	2.6	24.8	52.1	23.1	4.4	2.4	93.2
	(0.01)	(0.00)	(0.01)	(0.06)	(0.07)	(0.06)	(0.01)	(0.01)	(0.01)
Corrected									
Restrictive	98.2	0.9	0.9	12.0	79.7	8.3	1.3	1.4	97.3
	(0.01)	(0.01)	(0.01)	(0.09)	(0.12)	(0.07)	(0.01)	(0.01)	(0.01)
Baseline	93.0	3.4	3.6	37.5	40.2	22.3	7.0	3.9	89.1
	(0.10)	(0.08)	(0.06)	(1.02)	(0.82)	(0.47)	(0.12)	(0.08)	(0.13)
ICE	98.1	0.8	1.1	10.2	82.2	7.7	1.6	1.5	96.9
	(0.02)	(0.01)	(0.01)	(0.18)	(0.18)	(0.10)	(0.02)	(0.02)	(0.03)
Markov	97.1	1.6	1.4	19.0	68.7	12.3	2.5	2.0	95.4
	(0.04)	(0.03)	(0.03)	(0.32)	(0.38)	(0.19)	(0.05)	(0.03)	(0.06)
Stationarity	93.7	2.8	3.4	33.7	45.5	20.8	6.6	3.1	90.3
	(0.06)	(0.05)	(0.05)	(0.50)	(0.53)	(0.30)	(0.09)	(0.05)	(0.09)
More lags	92.8	3.3	3.9	32.1	40.3	27.6	6.1	5.0	88.9
	(0.23)	(0.18)	(0.21)	(1.52)	(1.94)	(1.35)	(0.45)	(0.29)	(0.35)

Note: Average transition probabilities over 1996–2019. In parentheses are bootstrapped standard errors based on 500 repetitions.