Journal of Asian Economics 75 (2021) 101319

Contents lists available at ScienceDirect

Journal of Asian Economics

Misclassification errors in labor force statuses and the early identification of economic recessions☆

Jiandong Sun[a], Shuaizhang Feng[a,b,*], Yingyao Hu[c]

[a] Institute for Economic and Social Research, Jinan University, Guangzhou 510632, China
[b] School of Economics, Jinan University, Guangzhou 510632, China
[c] Department of Economics, Johns Hopkins University, Baltimore, MD 21218, United States

ARTICLE INFO

Article history:
Received 4 March 2021
Received in revised form 13 April 2021
Accepted 16 April 2021
Available online 28 April 2021

Keywords:
Economic recession
Sahm rule
Misclassification errors
Unemployment rate

ABSTRACT

Accurate identification of economic recessions in a timely fashion is a major macroeconomic challenge. The so-called Sahm recession indicator (Sahm, 2019), one of the most reliable early detectors of the U.S. recessions, relies on changes in unemployment rates, and is thus subject to misclassification errors in labor force statuses based on survey data. We propose a novel misclassification-errors correction to improve the predictive timeliness and provide a proper threshold value. Using historical data, we show that the adjusted unemployment-based recession indicator offers earlier identification of economic recessions.

1. Introduction

The novel coronavirus (Covid-19) pandemic has been ripping through America not just with skyrocketing numbers of confirmed cases and deaths, but also its disruptive power on the U.S. economy. In June 2020, the Business Cycle Dating Committee of the National Bureau of Economic Research (NBER), which officially declares the turning points (peak and trough) of the U.S. business cycle, announced that a peak in monthly economic activity occurred in February 2020, implying that the U.S. economy has been in a recession.[1] Since the NBER's approach to determining the dates of business cycle turning points is retrospective, it would usually take the NBER several months to identify a recession after it has already occurred. Considering that recessions would bring enormous damage to the economy, it is important for policymakers to initiate prompt and efficient monetary policies or fiscal stimulus to reduce its negative effects. Although the NBER's procedure guarantees relatively precise dates of recessions, it is too slow for policy-making.

To be sure, researchers have been always invoking sophisticated econometrics methods and using economic data to predict economic activities and identify recessions as soon as they can.[2] Recently, Sahm (2019) proposed a so-called "Sahm recession indicator" to identify the U.S. recession, which is based on changes in unemployment rate. That is, if the three-months moving average of the national unemployment rate (U-3) rises 0.5 percentage points or more relative to its low during the previous 12 months, then the U.S. economy is already in a recession.[3] This unemployment-based recession

☆ Feng's research was supported by National Science Fund for Distinguished Young Scholars (Project Number: 71425005).
* Corresponding author.
E-mail addresses: jiandongsun11@gmail.com (J. Sun), shuaizhang.feng@foxmail.com (S. Feng), yhu@jhu.edu (Y. Hu).
[1] See https://www.nber.org/news/business-cycle-dating-committee-announcement-june-8-2020.
[2] For detailed literature review, refer to Section 2.
[3] See also at https://fred.stlouisfed.org/series/SAHMREALTIME.

http://dx.doi.org/10.1016/j.asieco.2021.101319

indicator correctly signals a recession in 2–4 months after it actually occurred since 1980, which represents a significant time-saving compared to the NBER. This indicator also compares favorably with other prediction methodologies in terms of accuracy and timeliness, as well as simplicity.

However, the official unemployment rate used to calculate the Sahm recession indicator is subject to the well-known issue of misclassification in labor force status (LFS) (Abowd & Zellner, 1985; Poterba & Summers, 1986). Feng and Hu (2013) show that ignoring misclassification errors in LFS leads to substantial underestimation of the unemployment rate. More importantly, they show that the degree of underestimation is larger when the level of unemployment is higher. In this sense, it is possible that at the very beginning of a recession, the rise in the official unemployment rate, which is subject to misclassification errors, is less than the increase in the underlying true unemployment rate. This may delay the signal of recessions and weaken the predictive timeliness of the unemployment-based recession indicator.

In this paper, we examine the robustness of the unemployment-based recession indicator to misclassification errors in LFS. Previous studies have widely discussed the issue of misclassification errors in LFS, which arises from the intrinsic difficulties in classifying LFS of some specific groups of people, like marginally-attached worker and involuntary part-time workers, as well as other practical challenges in classifying LFS with survey data. Using a latent variable approach, Feng and Hu (2013) correct for misclassification errors and estimate the "true" unemployment rate. We apply their method and use the corrected unemployment statistics to re-calculate the unemployment-based recession indicator. We find that for the historical recessions since the late 1970s, our adjusted recession indicator always rises more promptly than the original one after the onset of a recession. We also propose optimal threshold values for the identification of recessions and show that the misclassification-errors correction improves predictive timeliness of the unemployment-based recession indicator.

The rest of the paper is organized as follows. The next section reviews the literature related to identification of economic recessions and misclassification in labor force status. Section 3 briefly presents the misclassification model and discusses the assumptions. Section 4 presents the data and empirical results. The last section concludes.

2. Literature review

2.1. Identifying economic recessions

An economic recession typically occurs when there is a significant decline in economic activity for at least two consecutive quarters in a designated region. In the U.S., the NBER defines a recession as a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in production, employment, consumption and other indicators, among which the first two are the primary conceptual measures of economic activity. Since some indicators are only available quarterly and subject to substantial revisions and measurement errors, the NBER refers to a set of monthly indicators to determine the dates of turning points. On the one hand, the NBER normally views the monthly payroll employment data from the Current Employment Statistics (CES) as the most reliable comprehensive estimate of employment, with the household employment data from the Current Population Survey (CPS) as an alternative indicator. On the other hand, the NBER uses real personal income less transfer payments from the Bureau of Economic Analysis (BEA) as a monthly measure of output, as the most reliable measures of output, real gross domestic production (GDP) and gross domestic income (GDI), are only available quarterly.[4] To avoid major revisions, the NBER would wait for sufficient data to be available to determine the dates of turning points, leading to the delay of several months for identifying a recession after its occurrence.[5]

Researches have been trying to predict real economic activities and identify recessions using asset prices.[6] In particular, the yield spread, which refers to the difference in return between long-term and short-term government bonds, has been mostly frequently used in this endeavor.[7] The yield curve is normally upward sloped because the return of long-term bonds rates are higher than that of short-term ones. However, it is observed that the yield curve tends to be flat or even inverted nearly before recessions. Estrella and Mishkin (1998) find that the yield spread predicts recessions beyond one quarter.[8] Expectation theory explains that, since long-term interest rate reflects financial markets' expectations of future short-term interest rate, a flat or inverted yield spread might indicate that the market expects that future interest rate should fall when weak economic activity or a future recession is anticipated.[9] Other leading predictors include interest rate spreads (Stock & Watson, 1989; Wright, 2006), GDP and GDI (Nalewaik, 2012), the Conference Board Leading Economic Index (Lahiri & Yang, 2015; Levanon, Manini, Ozyildirim, Schaitkin, & Tanchua, 2015) and so on. Nevertheless, the predictive power of financial indicators might produce more false triggers, partly due to their immediate responses to monetary policies (Sahm, 2019).

[4] See more information at https://www.nber.org/research/business-cycle-dating.
[5] See the NBER's historical announcements of recessions at https://www.nber.org/research/business-cycle-dating/business-cycle-dating-committee-announcements.
[6] See Stock and Watson (2003) for more detailed review.
[7] The power of yield spread to predict recession was first demonstrated by Estrella and Hardouvelis (1991), Harvey (1989), Stock and Watson (1989).
[8] The power of yield spread to predict economic activities or recession can also be found in Ang, Piazzesi, and Wei (2006), Benati and Goodhart (2008), Croushore and Marsten (2016), Estrella (2005), Rudebusch and Williams (2009).
[9] For more possible explanations, we refer to detailed review in Tian and Shen (2019).

Methodologically, most existing studies focus on forecasting either output growth with continuous models, or a binary indicator of recession with binary models. Estrella, Rodrigues, and Schich (2003) find that binary models are more stable than continuous models. Considering the existence of autocorrelation in binary recession indicator, recent studies extend the standard static probit model with a dynamic setting, such as Kauppi and Saikkonen (2008) and Chauvet and Potter (2005), among others. Since the seminal work of Hamilton (1989), some studies employ Markov switching model to forecast recessions, such as Chauvet (1998), Chauvet and Hamilton (2006). Some studies show that the Markovian models outperform the other methods, such as nonparametric algorithms and probit models, in detecting a recession and capturing the recession duration (Chauvet & Piger, 2008; Tian & Shen, 2019).

In addition, in light of the need of timely policy decisions, there has been increasing interest in "nowcasting", that is, assessing or predicting current and very near future economic conditions. Giannone, Reichlin, and Small (2008) develop a formal method to nowcast real GDP growth. Hamilton (2011) surveys efforts to automate the dating of turning points of business cycles. Chauvet and Piger (2008) evaluate the ability of existing models to establish business cycle turning point dates in real time. Chen, So, Wu, and Yan (2015) use the data from Google search to improve the timeliness of the identification of business cycle turning points and nowcast the peak dates within the month that the turning occurred.

More recently, Sahm (2019) proposes that an automatic stimulus payments to individual should be triggered if the three-months moving average of the national unemployment rate (U-3) rises 0.5 percentage points or more relative to its low during the previous 12 months, which is the so-called "Sahm Rule" for identifying recessions. Instead of a fixed threshold of unemployment level, this rule uses changes in unemployment rate, which is a timely and consistent measure of labor market slack, to capture the signal of recessions. The smoothing procedure also rules out some monthly random variations to avoid false triggers. As a result, the unemployment-based recession indicator correctly signals a recession in 2–4 months after it actually occurred since 1980.

2.2. Correcting for misclassification in labor force status

Labor market statistics provide a useful way to understanding and monitoring the labor market, such as unemployment rate, labor force participation rate, gross labor flows and so on. They help to identify labor market slack, and lend support to the implementation of economic programs and subsequent evaluation, such as fiscal stimulus, training programs and unemployment insurance policies. In order to derive these labor market statistics, labor force surveys are therefore designed and conducted, and have been the primary source used to produce labor market statistics for several decades worldwide. According to the International Labor Organization (ILO)'s definition, individuals covered in the labor force surveys are generally classified into three statuses: "employed (E)", "unemployed (U)" and "not-in-labor-force (N)". However, the reliability of labor market statistics is highly dependent on the accuracy of the responses. Many sampling and non-sampling errors could bias these statistics. Misclassification across different LFS is one of the most well-observed issues. It occurs when respondents misreport some particular labor force items deliberately or misunderstand questions, or interviewers record answers incorrectly.

For example, due to the Covid-19 pandemic, the U.S. labor force statistics since March 2020 are subject to serious misclassification errors. As is denoted in the Employment Situation (March 2020), although special instructions have been given to household survey interviewers that all employed persons absent from work due to coronavirus-related business closures should be classified as unemployed on temporary layoff, it is apparent that not all such workers were so classified, because there were large increases in the number of persons who were classified as employed but absent from work for other reasons as well, as shown in Fig. 3.[10] The report further predicts:

> If the workers who were recorded as employed but absent from work due to "other reasons" (over and above the number absent for other reasons in a typical March) had been classified as unemployed on temporary layoff, the overall unemployment rate would have been almost 1 percentage point higher than reported.

Researchers have tried to use rigorous statistical models to identify misclassification errors in labor force status, irrespective of their sources, and to estimate the "true" labor force statistics. This strand of literature relies on the local independence assumption that misclassification errors in one period are unrelated to the errors in other periods conditional on the current true status, under which the misreporting behavior can be summarized as a simple 3-by-3 misclassification matrix as follows:

$$M_{S_t|S_t^*} = \begin{bmatrix} f_{S_t|S_t^*}(E|E) & f_{S_t|S_t^*}(E|U) & f_{S_t|S_t^*}(E|N) \\ f_{S_t|S_t^*}(U|E) & f_{S_t|S_t^*}(U|U) & f_{S_t|S_t^*}(U|N) \\ f_{S_t|S_t^*}(N|E) & f_{S_t|S_t^*}(N|U) & f_{S_t|S_t^*}(N|N) \end{bmatrix},$$

where $f(\cdot)$ stands for the probability mass functions of its arguments, S_t and S_t^* are reported and latent true LFS in period t, respectively.

[10] See more details at https://www.bls.gov/news.release/archives/empsit_04032020.htm.

Table 1
Misclassification probabilities in existing studies, Pr $(S_t = i|S_t^* = j)$.

	j = 1	j = 2	j = 3	j = 1	j = 2	j = 3	j = 1	j = 2	j = 3	j = 1	j = 2	j = 3
	AZ			PS			HMM			FH		
$i = 1$	98.8	1.9	0.5	97.8	3.8	1.2	98.7	7.5	1.4	97.9	17.3	2.9
$i = 2$	0.2	88.6	0.3	0.5	84.7	0.6	0.4	76.4	1.0	0.6	62.5	0.2
$i = 3$	1.0	9.5	99.2	1.7	11.5	98.2	0.9	16.1	97.6	1.5	20.2	96.9

Note: "AZ" refers to estimates by Abowd and Zellner (1985). "PS" refers to estimates by Poterba and Summers (1986). "HMM" refers to estimates by Shibata (2019). "FH" refers to estimates by Feng and Hu (2013).

To identify such a misclassification matrix, early studies use the reinterview surveys, in which a small subsample of original respondents were recontacted and asked the same questions. The responses from the reconciled reinterview surveys are typically treated as the "true" LFS (Abowd & Zellner, 1985; Magnac & Visser, 1999; Poterba & Summers, 1986; Singh & Rao, 1995). Nevertheless, the reinterview surveys also suffer from misclassification errors due to many practical limitations. Sinclair and Gastwirth (1996) also show that the reinterview surveys might have more errors than the original CPS sample. Another shortcoming is that these misclassification probabilities are based on historical reinterview surveys, which were released a long time ago, so the use of these estimates relies on the time-invariant assumption (Elsby, Hobijn, & Şahin, 2015).

More recent studies, such as Biemer and Bushery (2000), Bassi, Padoan, and Trivellato (2012), Feng and Hu (2013), Shibata (2019), utilize the panel structure of labor force surveys and take a latent variables approach, in which the dynamic process of the latent true LFS is jointly modeled together with the misclassification process.[11] Both Biemer and Bushery (2000) and Shibata (2019) use a maximum likelihood approach with local identification, while Feng and Hu (2013) take a novel eigenvalue-eigenvector decomposition approach to establish a closed-form global identification. Feng and Hu's proposed estimation procedure also directly leads to the unique estimates without choosing proper initial values or using regular time-consuming optimization algorithms. In terms of the assumption imposed on the dynamics of latent true LFS, Feng and Hu (2013) relax the first-order Markovian assumption which is widely used in many studies, including Biemer and Bushery (2000) and Shibata (2019). This assumption might be too restrictive due to the presence of state dependence, serial correlation and unobserved heterogeneity (Hyslop, 1999).

Table 1 compares the misclassification probabilities reported in the existing studies. In each matrix, the biggest element in each column lies in the diagonal of the matrix, which implies that respondents are more likely to report their true status than otherwise. The biggest misclassification errors come from the unemployed individuals' being classified as either not-in-labor-force or employed. This is consistent with the findings that some marginally-attached people are behaviorally more similar to the unemployed than to the rest of those not-in-labor-force (Jones & Riddell, 1999), and many involuntary part-time workers are also observationally more similar to the unemployed (Farber, 1999). The main difference among these matrices is the magnitude of misclassification probabilities. In general, those obtained by latent variables approach (Feng & Hu, 2013; Shibata, 2019) have much higher misclassification errors than those obtained from the reconciled reinterview surveys (Abowd & Zellner, 1985; Poterba & Summers, 1986).

In addition, some studies try to address the potential misclassification between *U* and *N* through isolating the transition sequences that involve the reversals of transitions between *U* and *N*. In particular, they simply re-classify all the *UNU* as *UUU*, and *NUN* as *NNN* (Elsby et al., 2015; Farber & Valletta, 2015; Rothstein, 2011). Undoubtedly, such a re-classification approach could adjust some spurious transitions between *U* and *N*, but might also over-correct some genuine transitions. Ahn and Hamilton (2020) modify this procedure and re-classify *UNU* as *UUU* only if the third *U* is together with a duration of job search greater than 4 weeks. Nevertheless, this approach is *ad hoc* in nature and does not have rigorous empirical or theoretical basis.

3. Methods

3.1. Estimation of misclassification probabilities

We use the method proposed by Hu (2008) and used in Feng and Hu (2013) to correct for biases in unemployment rates due to misclassification errors.[12] According to the 4-8-4 rotational group structure of the Current Population Survey, suppose we observe an i.i.d. sample of the reported labor force status for three periods $\{S_{t+1}, S_t, S_{t-9}\}_i$ for individual *i*. For example, if S_t stands for one's LFS in January 2020, then S_{t+1} and S_{t-9} denote the LFS in February 2020 and in April 2019, respectively. Although each person appears eight times in CPS, we choose these three-months data for the following reasons: (i) we want the three months to be close enough to minimize sample attrition; (ii) we want the three months to cover the eight-months break in the 4-8-4 rotation structure to ensure that there are enough variations in the LFS; (iii) the assumption regarding the dynamics of the latent true LFS (Assumption 2 below) is more likely to be satisfied if we use the data reported nine months ago.

[11] For more discussions on the latent variables approach, we refer to Hu (2017).
[12] See Feng and Hu (2013) for more technical details.

We assume that the latent true LFS S_t^* has the same support as the reported LFS S_t as follows:

$$S_t = \begin{cases} 1, \text{employed} \\ 2, \text{unemployed} \\ 3, \text{ not in labor force} \end{cases}.$$

Let $\Omega_{\neq t}$ denote all the variables in all the periods except period t, i.e., $\Omega_{\neq t} = \{(S_\tau, S_\tau^*), \tau \neq t\}$. We assume that the misclassification errors distribution satisfies a local independence assumption as follows:

Assumption 1. $f(S_t|S_t^*, \Omega_{\neq t}) = f(S_t|S_t^*)$.This assumption implies that misclassification errors may be correlated with the true LFS, and correlated with all other variables only through the true LFS.

In addition, we simplify the dynamics of the latent true LFS as follows:

Assumption 2. $f(S_{t+1}^*|S_t^*, S_{t-9}^*) = f(S_{t+1}^*|S_t^*)$.

This assumption implies that the true LFS in period $t-9$ has no predictive power on the true LFS in period $t+1$ beyond the true LFS in the current period t, which is weaker than the widely-used first-order Markovian assumption (Biemer & Bushery, 2000; Shibata, 2019).

Under Assumptions 1 and 2, the relationship between the observed probabilities and the unobserved ones can be derived as follows:

$$f(S_{t+1}, S_t, S_{t-9}) = \sum_{S_t^*} f(S_{t+1}|S_t^*) f(S_t|S_t^*) f(S_t^*, S_{t-9}). \tag{1}$$

By integrating S_{t+1} out, we obtain

$$f(S_t, S_{t-9}) = \sum_{S_t^*} f(S_t|S_t^*) f(S_t^*, S_{t-9}). \tag{2}$$

We then use the identification results in Hu (2008) to show that all the unobserved probabilities on the right-hand side of Eq. (1) can be identified. Define $M_{S_t|S_t^*} \equiv \left[f_{S_t|S_t^*}(i|j)\right]_{i,j}$, $M_{S_t^*,S_{t-9}} \equiv \left[f_{S_t^*,S_{t-9}}(j,k)\right]_{j,k}$, $M_{1,S_t,S_{t-9}} \equiv \left[f_{S_{t+1},S_t,S_{t-9}}(1,i,k)\right]_{i,k}$, and $D_{1|S_t^*} \equiv diag\left[f_{S_{t+1}|S_t^*}(1|j)\right]_j$. We can show that Eqs. (1) and (2) are equivalent to

$$M_{1,S_t,S_{t-9}} = M_{S_t|S_t^*} D_{1|S_t^*} M_{S_t^*,S_{t-9}}, \tag{3}$$

and

$$M_{S_t,S_{t-9}} = M_{S_t|S_t^*} M_{S_t^*,S_{t-9}}. \tag{4}$$

To solve the unknown misclassification probabilities, we need the following technical assumption:

Assumption 3. Matrix $M_{S_t,S_{t-9}}$ is invertible.

This assumption is testable, as it is imposed on observed probabilities. Under Assumption 3, we can derive following equation by eliminating $M_{S_t^*,S_{t-9}}$ in Eqs. (3) and (4):

$$M_{1,S_t,S_{t-9}} M_{S_t,S_{t-9}}^{-1} = M_{S_t|S_t^*} D_{1|S_t^*} M_{S_t|S_t^*}^{-1}. \tag{5}$$

This implies that the observed matrix on the left-hand side of Eq. (5) has an eigenvalue-eigenvector decomposition on the right-hand side.

In order to identify a unique decomposition, we need the following two additional assumptions:

Assumption 4. $f_{S_{t+1}|S_t^*}(1|k)$ are different for a different k.

Assumption 5. $f_{S_t|S_t^*}(k|k) > f_{S_t|S_t^*}(j|k)$ for $j \neq k$.

Assumption 4 can be also tested directly, as $f_{S_{t+1}|S_t^*}(1|k)$ for $k \in \{1,2,3\}$ are eigenvalues of observed matrix $M_{1,S_t,S_{t-9}} M_{S_t,S_{t-9}}^{-1}$. Assumption 5 implies that people are more likely to report their true LFS than any other possible status, which is consistent with the results from the reinterview surveys (Poterba & Summers, 1986) and other validation studies reviewed in Bound, Brown, and Mathiowetz (2001). Assumptions 4 and 5 guarantee that the eigenvalues are distinctive and that the eigenvectors can be ordered by the value of true labor force status.

As Hu (2017) points out, estimates based on eigenvalue-eigenvector decomposition might be outside the $[0,1]$ range or even complex values with a finite sample. A remedy is to treat Eq. (4) as moment conditions and estimate unknown probabilities with suitable restrictions as follows:

$$(\hat{M}_{S_t|S_t^*}, \widehat{D}_{1|S_t^*}) = arg\ \min ||M_{1,S_t,S_{t-9}} M_{S_t,S_{t-9}}^{-1} M_{S_t|S_t^*} - M_{S_t|S_t^*} D_{1|S_t^*}|| \tag{6}$$

$$s.t.\quad \begin{array}{l}(1)\text{ each entry of } M_{S_t^*} \text{ is in } [0,1],\\ (2)\text{ each column sum of } M_{S_t|S_t^*} \text{ equals 1 and } D_{1|S_t^*} \text{ is diagonal},\\ (3)\ D_{1|S_t^*} \text{ and } M_{S_t|S_t^*} \text{ satisfy Assumptions 4 and 5, respectively.}\end{array}$$

where $||\cdot||$ is the $\mathcal{L}_2$ norm.

3.2. Estimation of unemployment rate and labor force participation rate

We then present the estimation procedure of correcting for the biases in labor market statistics due to misclassification errors. For labor stock statistics, such as unemployment rate and labor force participation rate, we firstly estimate the marginal distribution of the latent true LFS from the following equation:

$$f_{S_t} = \sum_{S_t^*} f_{S_t|S_t^*} f_{S_t^*}. \tag{7}$$

Eq. (7) can further be rewritten as the following matrix notation:

$$M_{S_t} = M_{S_t|S_t^*} M_{S_t^*}, \tag{8}$$

where $M_{S_t} = \left[f_{S_t}(E), f_{S_t}(U), f_{S_t}(N)\right]^T$ and $M_{S_t} = \left[f_{S_t^*}(E), f_{S_t^*}(U), f_{S_t^*}(N)\right]^T$. Suppose that $M_{S_t|S_t^*}$ is invertible, $M_{S_t^*}$ may be identified through inverting the matrix $M_{S_t|S_t^*}$:

$$M_{S_t^*} = M_{S_t|S_t^*}^{-1} M_{S_t}.$$

In order to avoid estimated probabilities outside $[0,1]$, we use Eq. (8) as moment conditions and estimate them as follows:

$$\widehat{M}_{S_t^*} = arg\ \min ||M_{S_t|S_t^*} M_{S_t^*} - M_{S_t}|| \tag{9}$$

$$s.t.\quad \text{each entry of } M_{S_t^*} \text{ is in } [0,1].$$

Therefore, we could calculate the unemployment rate

$$UR = \frac{\widehat{f}_{S_t^*}(U)}{\widehat{f}_{S_t^*}(E) + \widehat{f}_{S_t^*}(U)},$$

and the labor force participation rate

$$LFPR = \widehat{f}_{S_t^*}(E) + \widehat{f}_{S_t^*}(U).$$

3.3. Estimation of gross labor flows

In the basic procedure of adjusting gross labor flow, the probability of observing a flow (S_{t-1}, S_t) when the true flow is (S_{t-1}^*, S_t^*), is assumed to be the product of the misclassification probability of observing status S_{t-1} while the true status is S_{t-1}^* and the probability of observing status S_t while the true status is S_t^*.[13] Under this assumption, one can relate the observed flow proportion and true flow proportion with misclassification probabilities as follows:

$$f_{S_{t-1},S_t} = \sum_{S_{t-1}^*, S_t^*} f_{S_{t-1}|S_{t-1}^*} f_{S_t|S_t^*} f_{S_{t-1}^*, S_t^*}. \tag{10}$$

To be concrete, we define the following 3-by-3 matrix as the joint distribution of observed flow between two consecutive periods, i.e., $t-1$ and t,

$$M_{S_{t-1},S_t} = \begin{bmatrix} f_{S_{t-1},S_t}(E,E) & f_{S_{t-1},S_t}(U,E) & f_{S_{t-1},S_t}(N,E) \\ f_{S_{t-1},S_t}(E,U) & f_{S_{t-1},S_t}(U,U) & f_{S_{t-1},S_t}(N,U) \\ f_{S_{t-1},S_t}(E,N) & f_{S_{t-1},S_t}(U,N) & f_{S_{t-1},S_t}(N,N) \end{bmatrix},$$

[13] See also in Singh and Rao (1995). For additional procedures, such as correcting for margin errors and time aggregation bias, we refer to Poterba and Summers (1986), Shimer (2012), Elsby et al. (2015), among others.

and $M_{S^*_{t-1},S^*_t}$ for the corresponding matrix of underlying true flow. Suppose that the matrix $M_{S_t|S^*_t}$ is time-invariant, Eq. (10) can be rewritten as the following notation:

$$M_{S_{t-1},S_t} = M_{S_t|S^*_t} M_{S^*_{t-1},S^*_t} M^T_{S_t|S^*_t}, \tag{11}$$

Suppose that $M_{S_t|S^*_t}$ is invertible, one can identify the true transition probabilities as follows:

$$M_{S^*_{t-1},S^*_t} = M^{-1}_{S_t|S^*_t} M_{S_{t-1},S_t} (M^{-1}_{S_t|S^*_t})^T.$$

In practice, we could use Eq. (11) as moment conditions to solve the following optimization problem with suitable restrictions:

$$\widehat{M}_{S^*_{t-1},S^*_t} = arg\ \min||M_{S_t|S^*_t} M_{S^*_{t-1},S^*_t} M^T_{S_t|S^*_t} - M_{S_{t-1},S_t}|| \tag{12}$$

$$s.t.\ \text{each entry of}\ \ M_{S^*_{t-1},S^*_t}\ \text{is in}\ [0,1].$$

Finally, the transition probabilities can be estimated as follows:

$$\Pr(S^*_t = j|S^*_{t-1} = i) = \frac{\widehat{f}_{S^*_{t-1},S^*_t}(i,j)}{\sum_k \widehat{f}_{S^*_{t-1},S^*_t}(i,k)},$$

where $i,j,k \in \{E,U,N\}$.

4. Data

We use the public-use monthly Current Population Surveys datasets to estimate misclassification probabilities. Because of the 4-8-4 rotational group structure, the monthly CPS files can be matched to form longitudinal panels, which enables us to obtain the joint distribution of reported LFS in three periods. We use the following algorithm to match the monthly CPS files (Feng, 2001, 2004, 2008, 2013; Madrian & Lefgren, 2000). We first match the CPS samples crudely using household and individual identifiers. Since there might be coding errors in identifiers, we further use demographic information on gender, age and race to ensure that the matched sample refers to the same individual.

It has been documented that the matched sample is not representative of the original cross-sectional sample due to sample attrition (Feng, 2008). In order to correct for attrition, we generate a matching weight. Specifically, we first run a Logit regression, where the dependent variable is a binary indicator, 1 for matched or 0 for not matched, and the independent variables include gender, race, age, education and the LFS. Second, we calculate the predicted probability of being matched for each observation in the matched sample. The final weight is the product of the original weight and the inverse of the predicted probability of being matched. This adjustment procedure ensures that the matched sample shares the same marginal distributions on the key individual characteristics with the original cross-sectional sample.

The distribution of reported LFS, M_{S_t}, is calculated as follows:

$$M_{S_t} = \begin{bmatrix} f_{S_t}(E) \\ f_{S_t}(U) \\ f_{S_t}(N) \end{bmatrix} = \begin{bmatrix} (1-UR_t) \times LFPR_t \\ UR_t \times LFPR_t \\ 1 - LFPR_t \end{bmatrix}.$$

UR_t and $LFPR_t$ are the seasonally-adjusted unemployment rate and labor force participation rate in period t, respectively, which are officially released by the U.S. Bureau of Labor Statistics (BLS).

5. Results

5.1. Corrected labor market statistics

In this subsection, we compare the corrected labor market statistics based on Feng and Hu's method with the reported ones, as well as the other estimates based on the misclassification matrix in Table 1. Fig. 1 shows the results for the unemployment rates during 1996–2018. On average, our corrected unemployment rate (the “FH” series) is 2.3 percentage points higher than the reported one. Although we do not report standard errors in the graph, the difference is statistically significant as shown in Feng and Hu (2013), implying a substantial underestimation of the U.S. official unemployment rate. In addition, the degree of underestimation is larger when the unemployment level is higher. In comparison, other corrected unemployment rates (the “AZ”, “PS” and “HMM” series) are also higher than the reported ones, with some exceptions in the “PS” series in a few periods. Typically, the estimates based on the latent variables approach (the “FH” and “HMM” series) are higher than those from the reinterview surveys (the “AZ” and “PS” series).

Fig. 2shows the results for the labor force participation rates during 1996–2018. Our corrected labor force participation rate (the “FH” series) is on average only 0.9 percentage points higher than the reported one, and is not statistically significant

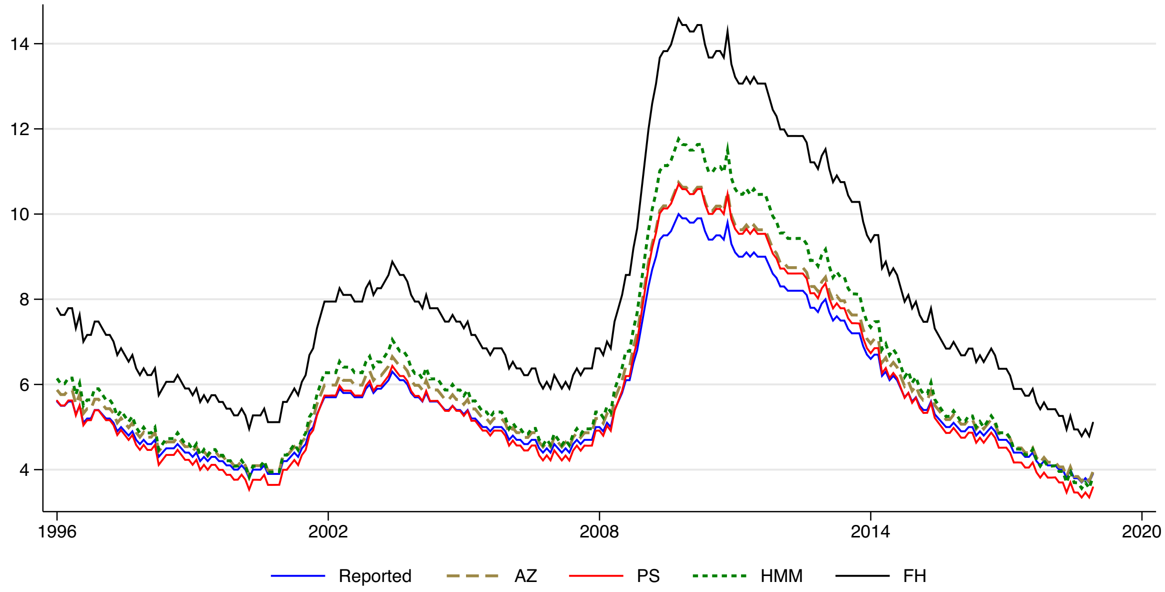

Fig. 1. Comparing the reported unemployment rates with the corrected ones in the existing studies, 1996–2018. *Note*: "Reported" refers to the official unemployment rates, while "AZ", "PS", "HMM" and "FH" refer to the corrected unemployment rates based on the misclassification matrix in Abowd and Zellner (1985), Poterba and Summers (1986), Shibata (2019) and Feng and Hu (2013), respectively.

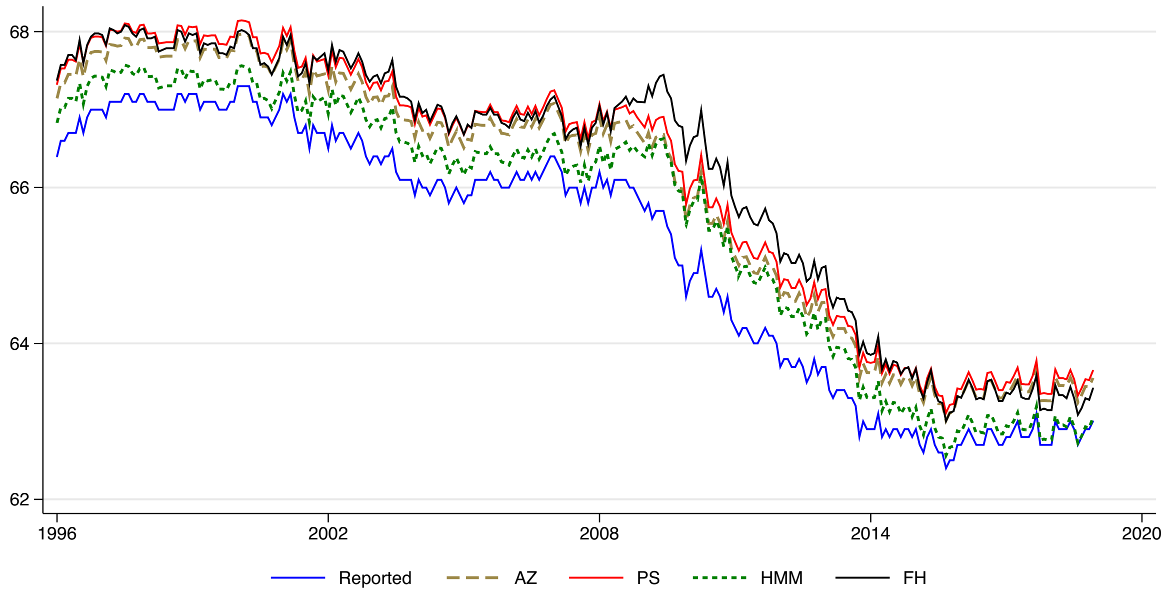

Fig. 2. Comparing the reported labor force participation rates with the corrected ones in the existing studies, 1996–2018. *Note*: "Reported" refers to the official labor force participation rates, while "AZ", "PS", "HMM" and "FH" refer to the corrected labor force participation rates based on the misclassification matrix in Abowd and Zellner (1985), Poterba and Summers (1986), Shibata (2019) and Feng and Hu (2013), respectively.

as shown in Feng and Hu (2013). This implies that misclassification errors may cause little bias to the labor force participation rate.

For labor flow statistics, Table 2 shows the average corrected transition probabilities over 1996–2018, based on the procedure presented in Section 3.3. Our corrected transition probabilities (the "FH" series) of staying in the same LFS as in the last period (i.e., the *EE*, *UU* and *NN* transitions) are higher than the reported ones, while all the corrected transitions across different LFS in two consecutive periods are lower than the reported ones. Such a direction of correction is also consistent with other estimates, though the magnitudes of these corrections are different. It is worth noting that, Shibata (2019) produces some implausible estimates in the "PS" and "FH" series without imposing the $[0, 1]$ restriction, but estimates the

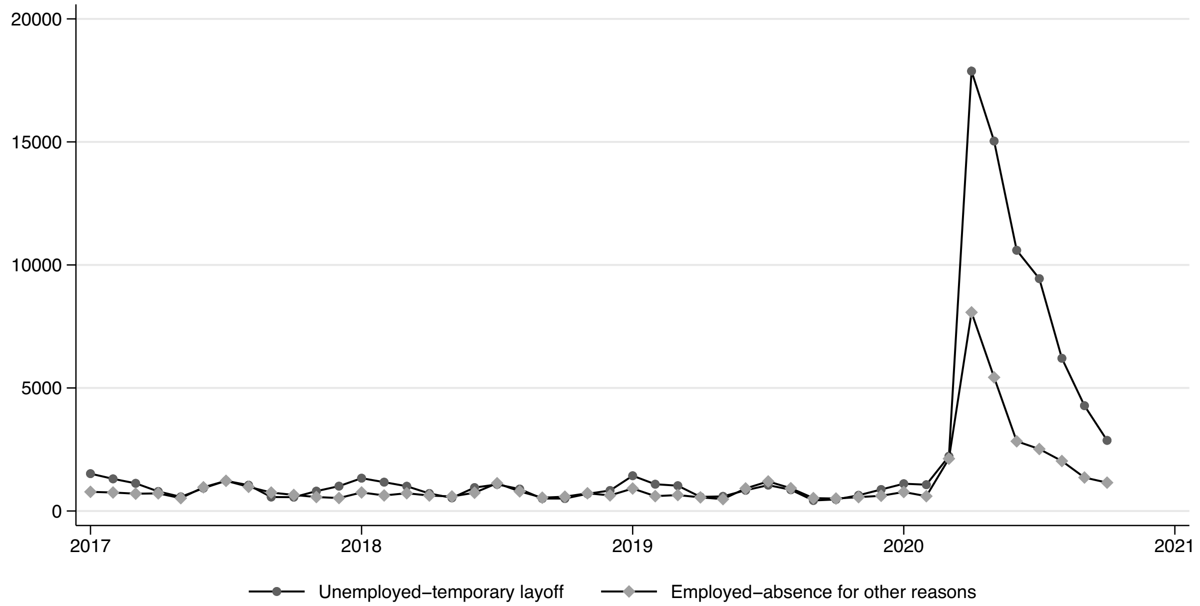

Fig. 3. The number of unemployed-temporary layoff vs. employed-absence for other reasons (number in thousands). *Source*: Authors' calculations from the Current Population Surveys 2017–2020. *Note*: Other reasons include reasons other than illness, vacation, weather, labor dispute, child care problems, other family or personal obligations, maternity or paternity leave, school or training, civic or military duty. Not seasonally adjusted.

Table 2
Average transition probabilities over 1996–2018, $f_{S_t|S_{t-1}}$ vs. $f_{S_t^*|S_{t-1}^*}$.

	$(E\|E)$	$(U\|E)$	$(N\|E)$	$(E\|U)$	$(U\|U)$	$(N\|U)$	$(E\|N)$	$(U\|N)$	$(N\|N)$
Reported	95.75	1.32	2.93	24.54	51.46	24.00	4.46	2.53	93.02
AZ	97.24	1.18	1.58	21.61	61.68	16.71	2.07	1.85	96.08
PS	98.72	0.78	0.51	16.02	70.71	13.27	0.27	1.37	98.37
HMM	98.10	0.77	1.14	15.35	79.94	4.71	1.20	0.47	98.33
FH	99.93	0.04	0.03	0.76	95.42	3.82	0.00	0.28	99.72

Note: "AZ", "PS", "HMM" and "FH" refer to the corrected transition probabilities based on the misclassification matrix in Abowd and Zellner (1985), Poterba and Summers (1986), Shibata (2019) and Feng and Hu (2013), respectively.

"HMM" series with such a restriction imposed under a first-order Markov assumption. Apparently, comparison among these estimates should not be used to infer which estimation method is better.

5.2. Adjusted unemployment-based recession indicator

In this subsection, we propose a misclassification-errors-adjusted unemployment-based recession indicator to identify recessions. The studying period is from January 1979 to October 2020, as there were no formal announcements of business cycle turning points prior 1979. Considering that we are focusing on the predictive timeliness of the recession indicator, we use real-time labor force statistics available in a given month.[14] In addition, when estimating misclassification probabilities, we pool different periods of matched data together to increase sample sizes. Specifically, the misclassification probabilities for period t, $M_{S_t|S_t^*}$, is estimated based on pooled matched samples from period $t-60$ to $t-1$, as in Feng, Hu, and Sun (2018).

Fig. 4 shows the seasonally-adjusted monthly reported and corrected unemployment rates, as well as the NBER-defined periods of economic recession. The reported values are directly from the BLS, and the corrected ones are calculated using the latent variable approach outlined in the previous sections. The shaded areas indicate economic recessions as per the definition of the NBER, which is the period between peak month (included) and subsequent trough month. It is clearly shown that during recessions, which generally post higher levels of unemployment, the differences between the corrected and

[14] Since the seasonal factors would be re-estimated at the end of each calendar year, the labor force statistics might be a little different at different vintages. See more details at https://www.bls.gov/web/empsit/cps-seas-adjustment-methodology.pdf. The labor force statistics at historical vintages can be retrieved from the Archival Federal Reserve Economic Data, Federal Reserve Bank of St. Louis. Note that the earliest vintage for labor force participation rate is 1997, so we use this vintage for labor force participation rate series during 1979–1996.

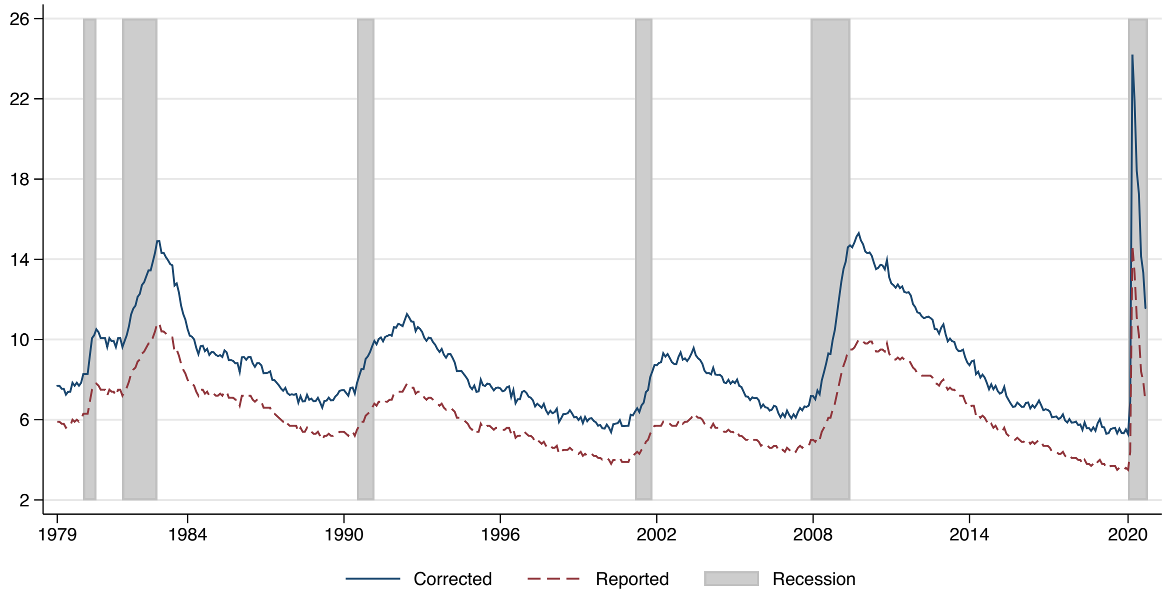

Fig. 4. Reported and corrected unemployemnt rates. *Source*: The U.S. Bureau of Labor Statistics and authors' calculations using the Current Population Surveys. *Note*: The reported and corrected unemployment rates are based on current vintage. All the series are seasonally adjusted.

reported unemployment rates are also much bigger than otherwise. This suggests that rises in the original Sahm recession indicator might have been suppressed and not truly reflecting changes in unemployment rates.

We then compare the real-time recession indicators based on both reported unemployment rates and our corrected ones in Fig. 5. During recessions, the indicator based on the corrected unemployment rate are higher than that based on the reported unemployment rate, while in expansions, the two indicators almost coincide. In addition, when recession is coming, the adjusted recession indicator based on our corrected unemployment rate always rises ahead of the original recession indicator.

We next turn to choose a threshold value for the adjusted unemployment-based recession indicator for identifying a recession. There is a trade-off in determining an optimal threshold, that is, the wrong triggers would increase with the threshold value in recessions but decrease in expansions. With the increase of threshold, we are more likely to reject false

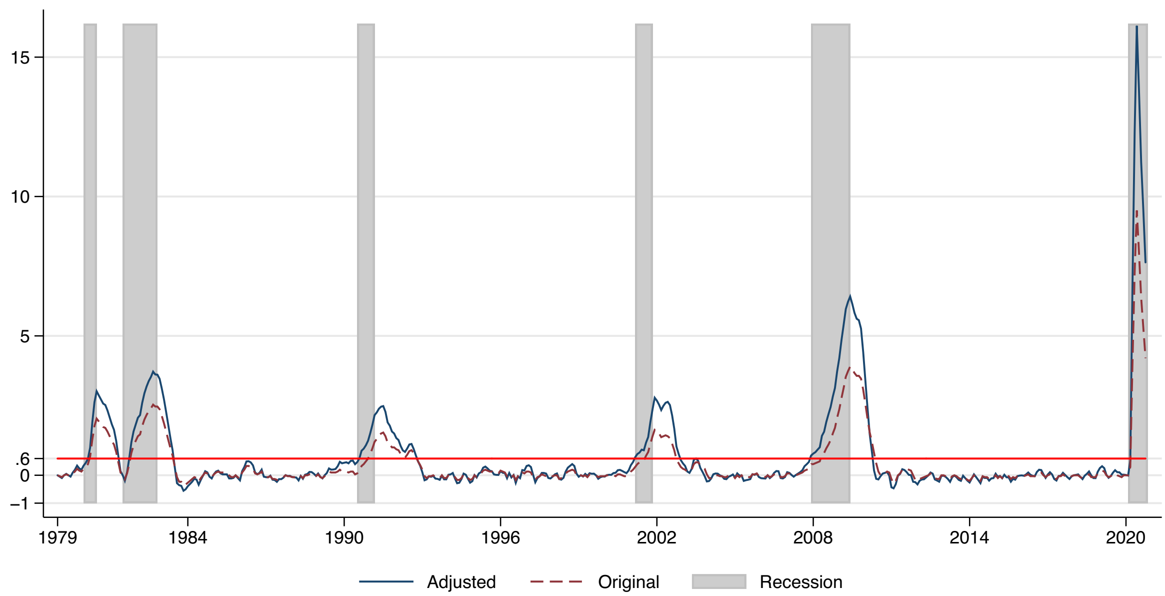

Fig. 5. Real-time Sahm recession indicator: original vs. adjusted. *Source*: The U.S. Bureau of Labor Statistics and authors' calculations from the Current Population Surveys. *Note*: The two recession indicators are produced using real-time data, that is, the seasonally-adjusted labor force statistics available in a given month.

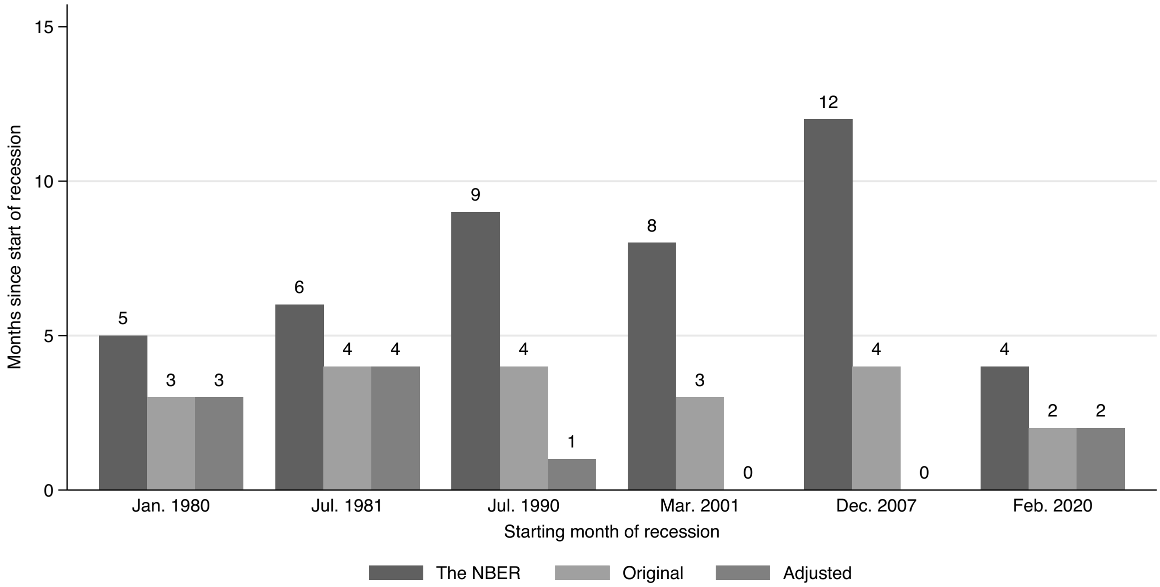

Fig. 6. Comparing the NBER's announcing dates with the triggering dates of Sahm recession indicators. *Source*: The National Bureau of Economic Research and authors' calculation.

alarms during the expansion periods, but are less likely to promptly identify the starting date of a recession when it comes. Taking into this consideration, we choose a value that is large enough to avoid false alarms. The idea is that we can tolerate some (unavoidable) delayed identification of true recessions but will exclude any false claim in order not to confuse the two. By this standard, we propose 0.6 as the threshold of the adjusted recession indicator, under which there would be no false alarm in expansions, as the original recession indicator performs.

Fig. 6 compares the timeliness of our adjusted unemployment-based recession indicator with the original one over the past U.S. recessions since 1980. In general, it would take the NBER from half a year to a full year to precisely identify the starting date of a recession. The original recession indicator identifies all the recessions within four months after they have already begun, which significantly improves upon the NBER. Our adjusted recession indicator outperforms the original one and substantially improves the timeliness of identification in half of the selected cases. For example, for the Great Recession that began in December 2007, the NBER announced it one year later in December 2008. The original indicator identifies the recessions in April 2008 with only four-months lag, while our adjusted indicator identifies it in December 2007, as soon as the household employment data for the month were released.

For the periods during the Covid-19 pandemic, the official U.S. unemployment rate has been jumping from 3.5% in February 2020, the lowest position in past 50 years, to 14.7% in April, the highest since the Great Depression era during the 1930s. After adjusting for misclassification errors, the corrected unemployment rate went from 5.3% in February, then 6.9% in March, to 24.2% in April, as shown in Fig. 4. Accordingly, the original Sahm recession indicator is 0.3 in March and 4.0 in April, while our adjusted one is 0.53 and 6.8, respectively, with both the original and our adjusted ones above their own cutoffs in April, as shown in Fig. 5. This implies that the U.S. economy has been already in a recession since April 2020, which is a two-months lag after the recession truly came, but two months ahead of the NBER's identification.

6. Conclusion

This paper examines the robustness of the Sahm recession indicator to misclassification errors in labor force statuses. Employing the latent variable approach used in Feng and Hu (2013), we correct for bias in unemployment rate due to misclassification errors, and re-calculate the recession indicator based on the corrected unemployment rate. We find that misclassification errors in labor force statuses do affect the predictive timeliness of the unemployment-based recession indicator. We then propose a more proper threshold for our adjusted recession indicator. Using historical records, our misclassification-errors correction substantially improves the predictive timeliness of the unemployment-based recession indicator.

References

Abowd, J. M., & Zellner, A. (1985). Estimating gross labor-force flows. *Journal of Business & Economic Statistics*, *3*(3), 254–283.
Ahn, H. J., & Hamilton, J. D. (2020). *Measuring labor-force participation and the incidence and duration of unemployment*. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 27394.
Ang, A., Piazzesi, M., & Wei, M. (2006). What does the yield curve tell us about GDP growth? *Journal of Econometrics*, *131*(1–2), 359–403.
Bassi, F., Padoan, A., & Trivellato, U. (2012). Inconsistencies in reported employment characteristics among employed stayers. *Statistica*, *72*(1), 93–109.

Benati, L., & Goodhart, C. (2008). Investigating time-variation in the marginal predictive power of the yield spread. *Journal of Economic Dynamics and Control*, *32*(4), 1236–1272.
Biemer, P. P., & Bushery, J. M. (2000). On the validity of Markov latent class analysis for estimating classification error in labor force data. *Survey Methodology*, *26*(2), 139–152.
Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In J. J. Heckman, & E. Leamer (Eds.), *Handbook of econometrics, vol. 5* (pp. 3705–3843). Elsevier.
Chauvet, M. (1998). An econometric characterization of business cycle dynamics with factor structure and regime switching. *International Economic Review*969–996.
Chauvet, M., & Hamilton, J. D. (2006). Dating business cycle turning points. In C. Milas, P. Rothman, & D. van Dijk (Eds.), *Nonlinear time series analysis of business cycles* (pp. 1–54). The Netherlands: Elsevier.
Chauvet, M., & Piger, J. (2008). A comparison of the real-time performance of business cycle dating methods. *Journal of Business & Economic Statistics*, *26*(1), 42–49.
Chauvet, M., & Potter, S. (2005). Forecasting recessions using the yield curve. *Journal of Forecasting*, *24*(2), 77–103.
Chen, T., So, E. P. K., Wu, L., & Yan, I. K. M. (2015). The 2007–2008 US recession: What did the real-time Google Trends data tell the United States? *Contemporary Economic Policy*, *33*(2), 395–403.
Croushore, D., & Marsten, K. (2016). Reassessing the relative power of the yield spread in forecasting recessions. *Journal of Applied Econometrics*, *31*(6), 1183–1191.
Elsby, M. W., Hobijn, B., & Şahin, A. (2015). On the importance of the participation margin for labor market fluctuations. *Journal of Monetary Economics*, *72*, 64–82.
Estrella, A. (2005). Why does the yield curve predict output and inflation? *The Economic Journal*, *115*(505), 722–744.
Estrella, A., & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *The Journal of Finance*, *46*(2), 555–576.
Estrella, A., & Mishkin, F. S. (1998). Predicting US recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, *80*(1), 45–61.
Estrella, A., Rodrigues, A. P., & Schich, S. (2003). How stable is the predictive power of the yield curve? Evidence from Germany and the United States. *Review of Economics and Statistics*, *85*(3), 629–644.
Farber, H. S. (1999). Alternative and part-time employment arrangements as a response to job loss. *Journal of Labor Economics*, *17*(S4), S142–S169.
Farber, H. S., & Valletta, R. G. (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, S. (2001). The longitudinal matching of current population surveys: A proposed algorithm. *Journal of Economic and Social Measurement*, *27*(1–2), 71–91.
Feng, S. (2004). Detecting errors in the current population survey: A matching approach. *Economics Letters*, *82*(2), 189–194.
Feng, S. (2008). Longitudinal matching of recent current population surveys: Methods, non-matches and mismatches. *Journal of Economic and Social Measurement*, *33*(4), 241–252.
Feng, S. (2013). Identification and statistical inference using matched march CPS data, with an application to us poverty dynamics. *Journal of Economic and Social Measurement*, *38*(2), 159–170.
Feng, S., & Hu, Y. (2013). Misclassification errors and the underestimation of the US unemployment rate. *The American Economic Review*, *103*(2), 1054–1070.
Feng, S., Hu, Y., & Sun, J. (2018). On the robustness of alternative unemployment measures. *Economics Letters*, *166*, 1–5.
Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676.
Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica,* 357–384.
Hamilton, J. D. (2011). Calling recessions in real time. *International Journal of Forecasting*, *27*(4), 1006–1026.
Harvey, C. R. (1989). Forecasts of economic growth from the bond and stock markets. *Financial Analysts Journal*, *45*(5), 38–45.
Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics*, *144*(1), 27–61.
Hu, Y. (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, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica*, *67* (6), 1255–1294.
Jones, S. R., & Riddell, W. C. (1999). The measurement of unemployment: An empirical approach. *Econometrica,* 147–161.
Kauppi, H., & Saikkonen, P. (2008). Predicting US recessions with dynamic binary response models. *The Review of Economics and Statistics*, *90*(4), 777–791.
Lahiri, K., & Yang, L. (2015). A further analysis of the conference board's new Leading Economic Index. *International Journal of Forecasting*, *31*(2), 446–453.
Levanon, G., Manini, J.-C., Ozyildirim, A., Schaitkin, B., & Tanchua, J. (2015). Using financial indicators to predict turning points in the business cycle: The case of the leading economic index for the United States. *International Journal of Forecasting*, *31*(2), 426–445.
Madrian, B. C., & Lefgren, L. J. (2000). An approach to longitudinally matching Current Population Survey (CPS) respondents. *Journal of Economic and Social Measurement*, *26*(1), 31–62.
Magnac, T., & Visser, M. (1999). Transition models with measurement errors. *Review of Economics and Statistics*, *81*(3), 466–474.
Nalewaik, J. J. (2012). Estimating probabilities of recession in real time using GDP and GDI. *Journal of Money, Credit and Banking*, *44*(1), 235–253.
Poterba, J. M., & Summers, L. H. (1986). Reporting errors and labor market dynamics. *Econometrica: Journal of the Econometric Society*1319–1338.
Rothstein, J. (2011). Unemployment insurance and job search in the great recession. *Brookings Papers on Economic Activity*, *2011*(Fall), 143–213.
Rudebusch, G. D., & Williams, J. C. (2009). Forecasting recessions: The puzzle of the enduring power of the yield curve. *Journal of Business & Economic Statistics*, *27*(4), 492–503.
Sahm, C. (2019). Direct stimulus payments to individuals. In H. Boushey, R. Nunn, & J. Shambaugh (Eds.), *Recession ready: Fiscal policies to stabilize the American economy* (pp. 67–92). Washington, DC: The Hamilton Project and the Washington Center on Equitable Growth.
Shibata, I. (2019). *Are labor market indicators telling the truth? Role of measurement error in the US Current Population Survey*. International Monetary Fund. Working Paper 19/40.
Shimer, R. (2012). Reassessing the ins and outs of unemployment. *Review of Economic Dynamics*, *15*(2), 127–148.
Sinclair, M. D., & Gastwirth, J. L. (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, A. C., & Rao, J. N. K. (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.
Stock, J. H., & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual*, *4*, 351–394.
Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, *41*(3), 788–829.
Tian, R., & Shen, G. (2019). Predictive power of Markovian models: Evidence from US recession forecasting. *Journal of Forecasting*, *38*(6), 525–551.
Wright, J. H. (2006). *The yield curve and predicting recessions*. Washington, DC: Federal Reserve Board. FEDs Working Paper 2006-7.