


# Rotation group bias and the persistence of misclassification errors in the Current Population Surveys

Shuaizhang Feng, Yingyao Hu & Jiandong Sun


To cite this article: Shuaizhang Feng, Yingyao Hu & Jiandong Sun (2022) Rotation group bias and the persistence of misclassification errors in the Current Population Surveys, *Econometric Reviews*, 41:9, 1077-1094, DOI: [10.1080/07474938.2022.2091361](https://doi.org/10.1080/07474938.2022.2091361)



To link to this article: <https://doi.org/10.1080/07474938.2022.2091361>

 View supplementary material 

 Published online: 12 Jul 2022.

 Submit your article to this journal 

 Article views: 143

 View related articles 

 View Crossmark data 



# Rotation group bias and the persistence of misclassification errors in the Current Population Surveys

Shuaizhang Feng<sup>a,b</sup>, Yingyao Hu<sup>c</sup>, and Jiandong Sun<sup>d</sup>

<sup>a</sup>Institute for Economic and Social Research, Jinan University, Guangzhou, China; <sup>b</sup>School of Economics, Jinan University, Guangzhou, China; <sup>c</sup>Department of Economics, Johns Hopkins University, Baltimore, Maryland, USA; <sup>d</sup>School of Economics and Statistics, Guangzhou University, Guangzhou, China

## ABSTRACT

We develop a general misclassification model to explain the so-called “Rotation Group Bias (RGB)” problem in the Current Population Surveys, where different rotation groups report different labor force statistics. The key insight is that responses to repeated questions in surveys can depend not only on unobserved true values, but also on previous responses to the same questions. Our method provides a framework to understand why unemployment rates in rotation group one are higher than those in other rotation groups in the CPS, without imposing any *a priori* assumptions on the existence and direction of RGB. Using our method, we provide new estimates of the U.S. unemployment rates, which are much higher than the official series, but lower than previous estimates that ignored persistence in misclassification.

## KEYWORDS

Current Population Survey; measurement error; panel data survey; rotation group bias

## JEL CLASSIFICATION

C14; E24; J21

## 1. Introduction

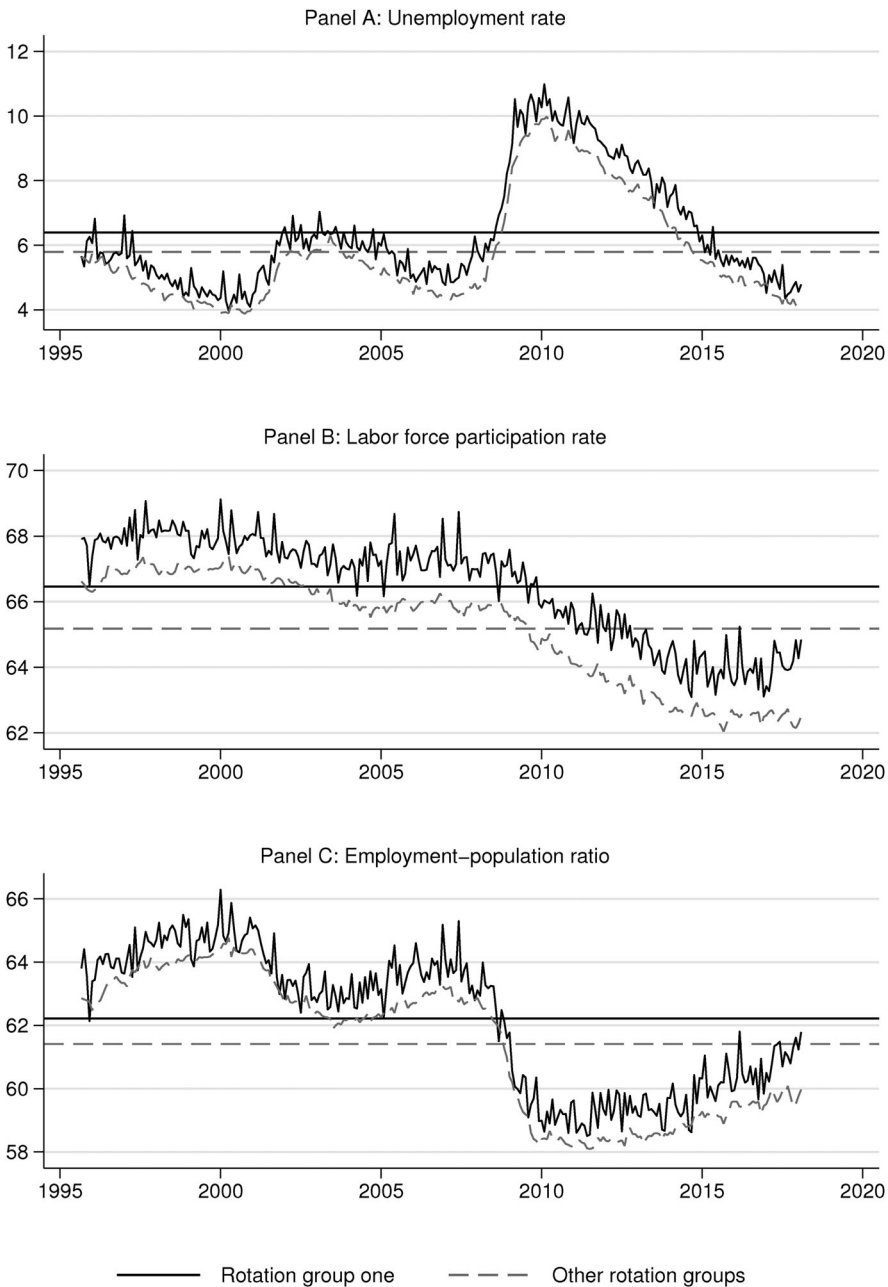
The Current Population Survey (CPS) has been the primary source of labor force statistics for the United States since the 1940s. A prominent feature of the CPS is its 4-8-4 rotating panel design, where an individual is interviewed for 4 consecutive months, idle for 8 months, and then interviewed for another 4 consecutive months before permanently being dropped out of the sample. In any given month, the CPS sample consists of eight rotation groups in the 4-8-4 rotating structure. In theory, all rotation groups should report the same statistics, apart from sampling errors, as they are all independent representative samples of the U.S. civilian population. However, it has been observed that rotation group one systematically posts much higher unemployment rates than other rotation groups, leading to the so-called phenomenon of “Rotation Group Bias (RGB)”.<sup>1</sup> As shown in Panel A of Fig. 1, the average unemployment rates over 1995-2018 for the first rotation group and others are 6.4% and 5.8%, respectively.

In this paper, we propose a measurement error model to explain the existence of RGB. We allow reported labor force status (LFS) to depend not only on latent true status, as in Feng and Hu (2013), but also on reported status in the previous interview. Rotation group one, and to a lesser extent, rotation group five, are structurally different from other rotation groups because they have no previously-reported status to condition on. Therefore, although reported

**CONTACT** Jiandong Sun ✉ [jiandongsun11@gmail.com](mailto:jiandongsun11@gmail.com) 📍 School of Economics and Statistics, Guangzhou University, 230 West Waihuan Road, Guangzhou 510006, China

📄 Supplemental data for this article is available online at <https://doi.org/10.1080/07474938.2022.2091361>.

<sup>1</sup>Historically, unemployment rates displayed a W-shaped pattern with rotation groups one and five posting higher levels. But the pattern has shifted to a monotonically decreasing one since the 1994 CPS redesign, as documented and carefully studied by Krueger et al. (2017).



**Figure 1.** Reported labor force statistics: rotation group one vs. other rotation groups.  
*Note:* Reported labor force statistics by rotation groups from September 1995 to February 2018. The horizontal lines are the corresponding average values. All numbers are seasonally-adjusted using the Census’s X-12-ARIMA algorithm.

unemployment rates from all rotation groups are potentially biased due to misclassification errors, there exist systematic differences in the (mis)reporting processes among rotation groups, leading to different reported statistics and hence RGB. Under certain conditional independence assumptions, we identify the distributions of misclassification errors and the distributions of latent true LFS for different rotation groups. To the best of our knowledge, this is the first rigorous statistical model incorporating RGB in a general misclassification framework.

Our model does not impose any assumptions regarding which rotation group would post more accurate statistics, nor does it have any *a priori* implications on the existence and direction of RGB. In contrast, our framework is flexible enough to allow for all possibilities regarding RGB.<sup>2</sup> This flexibility is important because not all rotating panel surveys display the same patterns of RGB as in the CPS, as noted by Krueger, Mas, and Niu (2017) for the Canadian Labor Force Survey.

We find strong evidence for our key assumption that the reported LFS depends not only on the latent true LFS, but also on the previously-reported LFS. For ongoing rotation groups where respondents were interviewed also in the immediate previous month, i.e., rotation groups 2–4 and 6–8, the misclassification matrices, which map true status into reported status, differ significantly for those who reported different LFS in the previous period. This suggests persistence in reporting behavior not captured by previous studies. Our out-of-sample corrections eliminate RGB between rotation group one and other rotation groups. We also provide estimated true unemployment rates, which are higher than the official ones.

Finally, in light of the evidence found in this paper, we examine the plausibility of alternative theories on panel conditioning (Halpern-Manners and Warren 2012). We rule out the “practice effect” hypothesis as our results suggest that later rotation groups underreport unemployment more than rotation group one. We show that the most likely explanations consistent with a higher unemployment rate for rotation group one are related to survey manipulations. Previous studies suggest that survey respondents may manipulate survey instruments in order to save time or energy costs associated with survey participation, or to avoid stigmatized answers such as being unemployed. We provide some empirical support for these hypotheses. However, the evidence is still preliminary and we do not rule out other mechanisms and alternative approaches to understanding RGB.<sup>3</sup>

The rest of the paper is organized as follows. Section 2 reviews prior studies related to RGB. Section 3 presents our measurement errors model and discusses key assumptions. Section 4 presents the main results, including estimated misclassification probabilities, out-of-sample corrections, and corrected labor force statistics based on our method. Section 5 discusses related studies and potential behavioral explanations for the observed RGB patterns. The last section concludes. The online Appendix includes data construction, detailed examinations of our key assumptions, robustness checks, and some additional results.

## 2. Literature review

Rotation group bias in the CPS has been well recognized and studied by researchers (Bailar 1975; Solon 1986; Erkens 2012; Krueger, Mas, and Niu 2017; Ahn and Hamilton 2022).<sup>4</sup> Many empirical exercises have been conducted to identify important factors that might affect rotation groups differently. A widely discussed factor is sample attrition, and in particular non-interviews, as rotation group one typically has higher Type-A non-interview rates than other rotation groups.<sup>5</sup> Although studies find that some RGB can be attributed to differences in response probabilities among rotation groups (Williams and Mallows 1970; Krueger, Mas, and Niu 2017), non-interviews do not explain the very existence of RGB. As Krueger, Mas, and Niu (2017) show, RGB

<sup>2</sup>In Section B of the online Appendix, we use Monte Carlo simulations to show that our framework can deal with positive, zero, and negative RGB.

<sup>3</sup>See more discussion in Section 5.

<sup>4</sup>RGB has also been observed in Canada (Ghangurde 1982), Netherlands (Van den Brakel and Krieg 2015) and New Zealand (Silverstone and Bell 2011).

<sup>5</sup>Type-A non-interview households are those eligible for the CPS interview but not interviewed because of refusal, temporary absence, non-contact, and other non-interview reasons.

exists even for those respondents who are surveyed in all eight interviews, and it remained stable before and after the 1994 CPS redesign.

Many other factors associated with the CPS interviewing process have also been discussed, including survey modes (computer-assisted vs. paper-and-pencil, telephone vs. in-person), types of survey respondents (self vs. proxy response), and types of interviewers (McCarthy 1978; U.S. Bureau of the Census 1978). These factors have been found to be inconsequential in general. For example, RGB remains even during the early CPS when all interviews were conducted in person (Hansen et al. 1955). In particular, the U.S. Census Bureau conducted a Methods Test Panel in 1978, including a set of twelve different treatment combinations in terms of survey modes, types of respondents, and types of interviewers.<sup>6</sup> Using the data from this experiment, Irvine (1984) finds that RGB is not attributable to any of these three types of treatments.

There is a strand of literature that studies how participating in a repeated survey may affect later responses to similar questions or subsequent behavior, which is known as “panel conditioning” (Warren and Halpern-Manners 2012). Different hypotheses have been proposed to explain why panel conditioning arises. On the one hand, panel conditioning might arise due to changes in true behavior. For example, respondents may become more conscious of the importance of the topics after being interviewed repeatedly, thereby changing their true behavior, known as the channel of “cognitive stimulus” (Sturgis, Allum, and Brunton-Smith 2009; Zwane et al. 2011). On the other hand, panel conditioning might also lead to changes in reporting behavior. It is possible that repeated interviewing may improve understanding of the rules that govern the interview process, leading to more accurate responses through the so-called “practice effect” (Bailar 1989; Waterton and Lievesley 1989). In contrast, repeated interviewing may also lead to inaccuracy. Respondents are more likely to report falsely to avoid socially non-normative or stigmatized answers, or to reduce the time or energy costs associated with survey participation, as noted by Warren and Halpern-Manners (2012). So far, no rigorous behavioral studies have been conducted with respect to the labor force status in the CPS to identify possible sources of RGB.

Lacking a clear understanding of why RGB arises, the Bureau of Labor Statistics report averages of all eight rotation groups in their official reports of labor force statistics. Other studies, such as Bailar (1975), focus instead on changes in unemployment rates, assuming RGB to be additive and constant over time. Nevertheless, Solon (1986) provides empirical tests that reject the assumption of additive rotation group effects. Alternatively, Ahn and Hamilton (2022) define RGB as a tendency that an individual changes his or her responses depending on the number of times surveyed, and use rotation group one as the benchmark to correct for RGB in other rotation groups. They also separately correct for general misclassification errors and nonrandom sample attrition common to all rotation groups.

On the other hand, there is a strand of statistical literature that tries to identify general measurement errors in the CPS, irrespective of their sources, thereby estimating the “true” labor force statistics. However, most of these studies have ignored RGB so far. These studies rely on the local independence assumption that the reported status in one period is independent of everything else conditional on the true status in the sample period, under which the misreporting behavior can be summarized as a simple 3-by-3 misclassification matrix. To identify such a misclassification matrix, some early studies use the responses from the reconciled reinterview surveys as the “true” LFS (Abowd and Zellner 1985; Poterba and Summers 1986, 1995; Magnac and Visser 1999). Other studies also use the reinterview data but impose strong assumptions for model identification (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. Utilizing the panel structure of the CPS, more recent studies develop a latent process of the underlying true LFS together with the misclassification process, such as Biemer and Bushery

---

<sup>6</sup>See details in (McCarthy 1978, Chap. 3).

(2000); Bassi and Trivellato (2009); Feng and Hu (2013); Shibata (2019). Nevertheless, the extent to which the bias in labor force statistics can be corrected depends on the specification of the misclassification errors. As Bound, Brown, and Mathiowetz (2001) point out, if measurement errors are persistent across periods, then the existing correction approaches might produce biased results.

Two previous studies are most similar to our paper in some methodological elements. Shockey (1988) study RGB in the CPS using a latent class approach, treating RGB as a result of different misclassification errors among rotation groups, with the same distribution of underlying LFS. However, Shockey (1988) only uses one-month CPS data and does not introduce persistence in misclassification errors. On the other hand, Bassi and Trivellato (2009) allow correlated classification errors by allowing the reported status in period  $t$  to depend on the true statuses both in periods  $t$  and  $t + 1$ , as their data is retrospective and information in period  $t$  is collected in period  $t + 1$ . In this paper, we extend existing models and introduce persistence in reporting behaviors by allowing reported status to not only depend on latent true status, but also on previously-reported status. Our approach takes care of RGB along with other misclassification errors common to all rotation groups simultaneously.

### 3. A model of rotation group bias

In this section, we develop a model based on the 4-8-4 rotating structure in the CPS. Suppose that for each individual in a random sample, we observe the reported LFS for two spells of four consecutive periods as follows:

$$\{U_1, U_2, U_3, U_4, U_{13}, U_{14}, U_{15}, U_{16}\}.$$

Between periods 4 and 13, there are 8 drop-out periods, on which we have no information as individuals are not interviewed. The reported LFS is defined as follows:

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

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

We are interested in how the reported status  $U_t$  is associated with the latent true status  $U_t^*$  in the whole process  $\{U_t, U_t^*\}$ . Following Hu (2008), we need four-periods matched data  $(U_{t+j}, U_t, U_{t-1}, U_{t-i})$  to identify misclassification probabilities. To maximize variations in LFS across different periods and to be consistent with our identification assumptions, we choose  $(U_{13}, U_4, U_3, U_1)$ .<sup>7</sup>

Let  $\Pr(\cdot)$  stand for the probability distribution function of its arguments, we outline the following assumption to describe the reporting behavior:

**Assumption 1.** *The reported status in period  $t$  only depends on the true status in period  $t$  and the reported status in period  $t - 1$ . That is,*

$$\Pr(U_t | U_t^*, \{U_s, U_s^*\}_{s \leq t-1}) = \Pr(U_t | U_t^*, U_{t-1}). \quad (1)$$

This assumption allows for correlation between the current reported status  $U_t$  and the previously-reported status  $U_{t-1}$  conditional on the current true status  $U_t^*$ , which can be supported by some

<sup>7</sup>In Section E of the online Appendix, we also provide robustness checks using alternative matched periods, such as  $(U_{13}, U_4, U_3, U_2)$  and  $(U_{13}, U_3, U_2, U_1)$ , and the results are still robust. The tradeoff is that if two periods are close to each other, the matrix  $M_{U_4, U_3, U_2}$  or  $M_{U_3, U_2, U_1}$  (required in Assumption 3) is more likely to be nearly singular, so the variances would be larger.

studies on panel conditioning, such as Halpern-Manners and Warren (2012).<sup>8,9</sup> Assumption 1 is much weaker than the widely-used local independence assumption, which further impose  $\Pr(U_t|U_t^*, U_{t-1}) = \Pr(U_t|U_t^*)$ , such as Feng and Hu (2013) among others.

Note that LFS is not observed in the immediate previous periods before the starting periods (periods 1 and 13). This implies that the misclassification errors may only depend on the current true value for the starting periods.<sup>10</sup> That is, for  $t = 1, 13$ ,

$$\Pr(U_t|U_t^*, \{U_s, U_s^*\}_{s \leq t-1}) = \Pr(U_t|U_t^*). \tag{2}$$

We then derive the joint distribution  $\Pr(U_{13}, U_4, U_3, U_1)$  as follows:

$$\begin{aligned} &\Pr(U_{13}, U_4, U_3, U_1) \\ &= \sum_{U_{13}^*, U_4^*} \Pr(U_{13}|U_{13}^*, U_4, U_4^*, U_3, U_1) \Pr(U_{13}^*|U_4, U_4^*, U_3, U_1) \Pr(U_4|U_4^*, U_3, U_1) \Pr(U_4^*, U_3, U_1) \\ &= \sum_{U_{13}^*, U_4^*} \Pr(U_{13}|U_{13}^*) \Pr(U_{13}^*|U_4, U_4^*, U_3, U_1) \Pr(U_4|U_4^*, U_3) \Pr(U_4^*, U_3, U_1). \end{aligned} \tag{3}$$

Having assumed the conditional independence of the reporting process, our next assumption deals with the dynamics of the latent true LFS:

**Assumption 2.** *Conditional on the true status in period  $t$  and the reported status in period  $t - 1$ , the true status nine months later is independent of the reported values in other periods. That is,*

$$\Pr(U_{t+9}^*|U_t^*, \{U_s\}_{s \leq t}) = \Pr(U_{t+9}^*|U_t^*, U_{t-1}). \tag{4}$$

Technically, this assumption is not completely comparable with that in Feng and Hu (2013), which assumes  $\Pr(U_{t+9}^*|U_{t+8}^*, U_{t-1}^*) = \Pr(U_{t+9}^*|U_{t+8}^*)$ , as our Assumption 2 involves reported status.<sup>11</sup> Nevertheless, they are both much weaker than the Markov process assumption that the whole process  $\{U_t, U_t^*\}$  is Markovian and  $\Pr(U_{t+1}^*|U_t^*, U_t) = \Pr(U_{t+1}^*|U_t^*)$ , such as Biemer and Bushery (2000); Shibata (2019).

Assumption 2 implies that

$$\Pr(U_{13}^*|U_4, U_4^*, U_3, U_1) = \Pr(U_{13}^*|U_4^*, U_3). \tag{5}$$

<sup>8</sup>In Section B of the online Appendix, we use Monte Carlo simulations to evaluate the robustness of this assumption through two deviations. The first is  $\Pr(U_t|U_t^*, \{U_s, U_s^*\}_{s \leq t-1}) = \Pr(U_t|U_t^*, U_{t-1}, U_{t-2})$ , and the second is  $\Pr(U_t|U_t^*, \{U_s, U_s^*\}_{s \leq t-1}) = \Pr(U_t|U_t^*, U_{t-1}, U_{t-1}^*)$ . The simulation results show that our estimators are robust to reasonable deviations.

<sup>9</sup>We have also tried to allow for higher-order persistence in the misclassification error process along the lines of Bollinger and David (2005), and estimated  $\Pr(U_t|U_t^*, U_{t-1}, U_{t-2})$  using the CPS data. However, this would require more periods of matched data, leading to large standard errors in some estimates. The results are similar to our baseline results, as shown in Section E of the online Appendix.

<sup>10</sup>For period 13, it is possible that people still remember their previous reports after eight drop-out periods. However, it is important to note that similar to period 1 (the first interview), dependent interviewing is also not used in period 13 (the fifth interviews), which might considerably weaken the correlations between reported value in period 13 and previous reports. Please refer to <https://www.census.gov/programs-surveys/cps/technical-documentation/methodology/collecting-data.html>. Despite this, we perform simulation study to show that even if this assumption is relaxed to  $\Pr(U_{13}|U_{13}^*, \{U_s, U_s^*\}_{s \leq 12}) = \Pr(U_{13}|U_{13}^*, U_4)$  or  $\Pr(U_t|U_t^*, \{U_s, U_s^*\}_{s \leq t-1}) = \Pr(U_t|U_t^*, U_{t-1}^*)$ , our proposed estimators are still robust. See Section B of the online Appendix.

<sup>11</sup>Suppose that there is no misclassification error and the  $U_{t-1}$  in  $\Pr(U_{t+9}^*|U_t^*, U_{t-1})$  is replaced with  $U_{t-1}^*$ , we may be able to argue that our Assumption 2 is weaker than that in Feng and Hu (2013), because our assumption is imposed on true labor dynamics across eight drop-out periods, while theirs is imposed across two consecutive periods.

<sup>12</sup>In Section B of the online Appendix, we use simulations and show that even if relaxing this assumption to  $\Pr(U_{13}^*|U_4, U_4^*, U_3, U_1) = \Pr(U_{13}^*|U_4^*, U_3, U_1)$  or  $\Pr(U_{13}^*|U_4, U_4^*, U_3, U_1) = \Pr(U_{13}^*|U_4^*, U_3, U_4)$ , our estimators are still robust.

Therefore, Equation (3) can be further written as follows:

$$\begin{aligned} & \Pr(U_{13}, U_4, U_3, U_1) \\ &= \sum_{U_4^*} \left( \sum_{U_{13}^*} \Pr(U_{13}|U_{13}^*)\Pr(U_{13}^*|U_4^*, U_3) \right) \Pr(U_4|U_4^*, U_3)\Pr(U_4^*, U_3, U_1) \\ &= \sum_{U_4^*} \Pr(U_{13}|U_4^*, U_3)\Pr(U_4|U_4^*, U_3)\Pr(U_4^*, U_3, U_1). \end{aligned} \tag{6}$$

For each value of  $U_3 = u_3$ , we have

$$\Pr(U_{13}, U_4, u_3, U_1) = \sum_{U_4^*} \Pr(U_{13}|U_4^*, u_3)\Pr(U_4|U_4^*, u_3)\Pr(U_4^*, u_3, U_1). \tag{7}$$

This means that we may apply the identification strategy in Hu (2008) to identify the unknown conditional distributions on the right hand side for each  $u_3$ , and therefore, to identify the misclassification probabilities  $\Pr(U_4|U_4^*, U_3)$ .

### 3.1. Identification of misclassification probabilities for ongoing rotation groups

In this subsection, we specify the main procedures to identify the misclassification probabilities for ongoing rotation groups, i.e., rotation groups 2–4 and 6–8. We start with Equation (7) with  $U_{13} = 1$  and  $U_3 = u_3$ .<sup>13</sup> That is,

$$\begin{aligned} & \Pr(U_{13} = 1, U_4, U_3 = u_3, U_1) \\ &= \sum_{U_4^*} \Pr(U_4|U_4^*, U_3 = u_3)\Pr(U_{13} = 1|U_4^*, U_3 = u_3)\Pr(U_4^*, U_3 = u_3, U_1). \end{aligned} \tag{8}$$

We first define matrices as follows:

$$\begin{aligned} M_{1, U_4, u_3, U_1} &= [\Pr(U_{13} = 1, U_4 = i, U_3 = u_3, U_1 = j)]_{i,j}, \\ M_{U_4, u_3, U_1} &= [\Pr(U_4 = i, U_3 = u_3, U_1 = j)]_{i,j}, \\ D_{1|U_4^*, u_3} &= \text{Diag}[\Pr(U_{13} = 1|U_4^* = j, U_3 = u_3)]_j, \\ M_{U_4|U_4^*, u_3} &= [\Pr(U_4 = i|U_4^* = j, U_3 = u_3)]_{i,j}, \\ M_{U_4^*, u_3, U_1} &= [\Pr(U_4^* = i, U_3 = u_3, U_1 = j)]_{i,j}. \end{aligned}$$

Therefore, Equation (8) is equivalent to

$$M_{1, U_4, u_3, U_1} = M_{U_4|U_4^*, u_3} D_{1|U_4^*, u_3} M_{U_4^*, u_3, U_1}. \tag{9}$$

Similarly, we have

$$M_{U_4, u_3, U_1} = M_{U_4|U_4^*, u_3} M_{U_4^*, u_3, U_1}. \tag{10}$$

In order to identify the unknown matrix  $M_{U_4|U_4^*, u_3}$ , we need a technical assumption as follows:

**Assumption 3.** For each  $u_3$ ,  $M_{U_4, u_3, U_1}$  has a full rank.

Since this assumption is imposed on the observed probabilities matrix, it can be directly tested using the CPS data. We use bootstrapping to show that the determinant of the observed matrix  $M_{U_4, u_3, U_1}$  for each  $u_3$  is significantly different from zero even if controlling for observed heterogeneity,

<sup>13</sup>The identification procedure holds for each value of  $U_{13}$ , implying that the model is over-identified. Here we use the subsample with  $U_{13} = 1$  to illustrate the identification strategy. In Section E of the online Appendix, we also provide robust results using different subsamples.



which means that  $M_{U_4, u_3, U_1}$  is invertible.<sup>14</sup> Eliminating  $M_{U_4^*, u_3, U_1}$  in Equation (9) and (10) leads to

$$M_{1, U_4, u_3, U_1} M_{U_4^*, u_3, U_1}^{-1} = M_{U_4|U_4^*, u_3} D_{1|U_4^*, u_3} M_{U_4^*|U_4^*, u_3}^{-1}. \tag{11}$$

Following Hu (2008),  $\Pr(U_4|U_4^*, U_3 = u_3)$  for each  $u_3$  can be identified from an eigen-decomposition.

In order to determine the eigenvector uniquely for each given eigenvalue, we need the eigenvalues to be distinctive, which can be formally stated as follows:

**Assumption 4.**  $\Pr(U_{13} = 1|U_4^* = i, U_3 = u_3) \neq \Pr(U_{13} = 1|U_4^* = j, U_3 = u_3)$  for  $i \neq j$ .

This assumption is also testable, as  $\Pr(U_{13} = 1|U_4^* = i, U_3 = u_3)$  for  $i = 1, 2, 3$  are the three eigenvalues of the observed matrix  $M_{u_{13}, U_4, u_3, U_1} M_{U_4^*, u_3, U_1}^{-1}$ . We can test this assumption directly by calculating the differences between the eigenvalues. The bootstrapping results show that the absolute differences between the eigenvalues are significantly different from zero even if controlling for observed heterogeneity, which means that the eigenvalues are distinctive. Intuitively, this assumption implies that the current true status has an impact on the probability of reporting being employed nine months later.

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

**Assumption 5.** Given  $U_3 = k$ , (i)  $\Pr(U_4 = k|U_4^* = k, U_3 = k)$  is the largest element in row  $k$ ; (ii) for  $i \neq j \neq k$ ,  $\Pr(U_4 = i|U_4^* = j, U_3 = k)$  is the smallest element in column  $j$ .

Assumption 5 implies that given the previously-reported status, when the 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, when the true status is different from the previously-reported status, then the least possible choice for the individual to report would be the status other than the true or previously-reported status.<sup>15</sup> Assumption 5 is an intuitive extension of the standard assumption in the literature that the reported status is assumed to only depend on the latent true status, and people are more likely to report the truth than otherwise, such as in Feng and Hu (2013) and other studies reviewed by Bound, Brown, and Mathiowetz (2001). In our extended framework, both the latent true status and the previously-reported status could matter in the current interview.

We summarize the closed-form identification and estimation of the misclassification probabilities  $\Pr(U_4|U_4^*, U_3)$  as follows:

**Proposition 1.** Under Assumptions 1 to 5, the misclassification probabilities  $\Pr(U_4|U_4^*, U_3)$  is uniquely determined by the observed joint probabilities of the reported LFS in four periods, i.e.,  $\Pr(U_{13}, U_4, U_3, U_1)$ , through the unique eigenvalue-eigenvector decomposition in Equation (11).

In fact, under the maintained assumptions above, we may use the method in Hu and Shum (2012) to identify the misclassification probabilities for each ongoing rotation group. For the convenience of estimation, we impose the stationarity and apply the misclassification probabilities  $\Pr(U_4|U_4^*, U_3)$  to all other ongoing rotation groups, instead of identifying them separately, i.e.,  $\Pr(U_\kappa|U_\kappa^*, U_{\kappa-1}) = \Pr(U_4|U_4^*, U_3)$  for  $\kappa = 2, 3, 14, 15, 16$ .

<sup>14</sup>See Section C of the online Appendix. The statistical inference is based on bootstrap. For example, we construct the bootstrap samples by resampling with replacement from the matched samples  $(U_{13}, U_4, U_3, U_1)$  for  $B$  repetitions. Then the bootstrap estimates of standard error can be calculated as follows:

$$s_{\hat{p}, B} = \sqrt{\frac{1}{B-1} \sum_{i=1}^B (\hat{p}_i - \bar{p})^2},$$

where  $\hat{p}_i$  is the bootstrap estimates and  $\bar{p} = \frac{1}{B} \sum_{i=1}^B \hat{p}_i$ .

<sup>15</sup>In Section D of the online Appendix, we illustrate how Assumption 5 works. Furthermore, we show results from alternative ordering possibilities and discuss the validity of this assumption.

### 3.2. Identification of misclassification probabilities for rotation groups one and five

To ensure that each rotation group posts the same true labor force statistics, we impose the following assumption similar to Shockey (1988):

**Assumption 6.** *In a given calendar month, the marginal distribution of the true LFS does not depend on rotation groups.*

As discussed in the last subsection, we may just use  $\Pr(U_4|U_4^*, U_3)$  as the misclassification probabilities for all ongoing rotation groups, and then achieve the distribution of the true LFS. Let  $\tau$  stand for a calendar month,  $\Pr(U_{i,\tau}^*)$  for  $i=1, 13$  be the distribution of the true LFS for rotation groups one and five respectively, and  $\Pr(U_{o,\tau}^*)$  be the distribution of the true LFS for ongoing rotation groups with  $o \in \{2, 3, 4, 14, 15, 16\}$  in month  $\tau$ . In addition, we define

$$p(U_\tau^*) = \begin{bmatrix} \Pr(U_\tau^* = 1) \\ \Pr(U_\tau^* = 2) \\ \Pr(U_\tau^* = 3) \end{bmatrix},$$

and

$$p(U_\tau^*, u_{\tau-1}) = \begin{bmatrix} \Pr(U_\tau^* = 1, U_{\tau-1} = u_{\tau-1}) \\ \Pr(U_\tau^* = 2, U_{\tau-1} = u_{\tau-1}) \\ \Pr(U_\tau^* = 3, U_{\tau-1} = u_{\tau-1}) \end{bmatrix}.$$

Therefore,  $p(U_{o,\tau}^*)$  for each  $\tau \in \{t_1, t_2, \dots, t_n\}$  can be estimated as follows:

$$p(U_{o,\tau}^*) = \sum_{j=1}^3 p(U_{o,\tau}^*, u_{o,\tau-1} = j) = \sum_{j=1}^3 M_{U_4|U_4^*, u_3=j}^{-1} \times p(U_{o,\tau}, u_{o,\tau-1} = j).$$

Assumption 6 implies that

$$\left[ p(U_{o,t_1}^*), p(U_{o,t_2}^*), \dots, p(U_{o,t_n}^*) \right] = \left[ p(U_{i,t_1}^*), p(U_{i,t_2}^*), \dots, p(U_{i,t_n}^*) \right],$$

for  $i=1, 13$ . Since we also observe the distribution of the reported LFS for rotation groups one and five, i.e.,  $[p(U_{i,t_1}), p(U_{i,t_2}), \dots, p(U_{i,t_n})]$  for  $i=1, 13$ , thus we have the matrix form with the stationarity as follows:

$$\begin{aligned} & [p(U_{i,t_1}), p(U_{i,t_2}), \dots, p(U_{i,t_n})] \\ &= M_{U_i|U_i^*} \left[ p(U_{i,t_1}^*), p(U_{i,t_2}^*), \dots, p(U_{i,t_n}^*) \right] \\ &= M_{U_i|U_i^*} \left[ p(U_{o,t_1}^*), p(U_{o,t_2}^*), \dots, p(U_{o,t_n}^*) \right]. \end{aligned} \tag{12}$$

The only unknown in Equation (12) is the misclassification matrix in the starting periods,  $M_{U_i|U_i^*}$ . We may identify and estimate  $M_{U_i|U_i^*}$  with observations from enough number of calendar months. In practice, we estimate  $M_{U_i|U_i^*}$  by minimizing the  $\mathcal{L}_2$  distance between the left-hand side and the right-hand side of Equation (12).<sup>16</sup> Since we only deal with sample proportions in different subpopulations, the consistency and the asymptotic properties can be straightforwardly derived and are omitted here.

---

<sup>16</sup>In estimation, we impose the restriction that all probabilities are in  $[0, 1]$ , see Hu (2017) for details.

**Table 1.** Misclassification probabilities.

	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
Panel A: Misclassification probabilities, $\Pr(U_4 = i   U_4^* = j, U_3)$									
	$U_3 = 1$			$U_3 = 2$			$U_3 = 3$		
$i=1$	<b>98.8</b> (0.02)	74.0 (0.73)	55.9 (0.46)	48.5 (0.52)	3.3 (0.45)	<b>7.4</b> (0.32)	41.7 (0.31)	<b>5.9</b> (0.39)	0.8 (0.02)
$i=2$	0.5 (0.02)	19.7 (0.55)	<b>2.2</b> (0.34)	41.5 (0.57)	<b>96.3</b> (1.26)	41.4 (1.41)	<b>6.1</b> (0.21)	38.9 (0.96)	0.4 (0.03)
$i=3$	0.7 (0.01)	<b>6.3</b> (0.66)	41.9 (0.65)	<b>9.9</b> (0.31)	0.3 (1.05)	51.2 (1.29)	52.2 (0.28)	55.2 (0.95)	<b>98.8</b> (0.04)
Panel B: Counterfactual misclassification probabilities, $\Pr(U_4 = i   U_4^* = j, \tilde{U}_3)$									
	$\tilde{U}_3 = 1$			$\tilde{U}_3 = 2$			$\tilde{U}_3 = 3$		
$i=1$	96.8 (0.02)	18.9 (0.27)	2.3 (0.02)	96.8 (0.08)	17.9 (1.03)	2.3 (0.10)	96.8 (0.03)	19.6 (0.35)	2.3 (0.03)
$i=2$	1.1 (0.01)	56.6 (0.31)	0.7 (0.04)	1.1 (0.06)	56.3 (1.32)	0.6 (0.16)	1.0 (0.02)	56.1 (0.43)	0.6 (0.05)
$i=3$	2.1 (0.01)	24.5 (0.36)	97.0 (0.05)	2.1 (0.06)	25.8 (1.46)	97.2 (0.21)	2.1 (0.02)	24.3 (0.49)	97.1 (0.07)
Panel C: Misclassification probabilities, $\Pr(U_5 = i   U_5^* = j)$									
	RG one			RG five					
$i=1$	95.1 (0.57)	25.9 (2.16)	9.5 (1.09)	97.1 (0.60)	29.2 (2.10)	3.0 (1.13)			
$i=2$	0.0 (0.00)	61.0 (1.42)	3.8 (0.19)	0.0 (0.00)	57.1 (1.42)	3.1 (0.19)			
$i=3$	4.9 (0.57)	13.1 (1.99)	86.7 (1.09)	3.0 (0.60)	13.7 (2.09)	94.0 (1.14)			

Note: In parentheses are bootstrapped standard errors based on 500 repetitions.

## 4. Empirical results

### 4.1. Misclassification probabilities

We use the public-use monthly CPS data from September 1995 to February 2018. Figure 1 displays the reported unemployment rates, labor force participation rates and employment-population ratios. There is clear evidence of RGB, that is, rotation group one posts higher levels in almost all periods among the three labor force statistics. These differences are also statistically significant although we do not report standard errors in the graph. On average, the reported unemployment rate from rotation group one is about 0.61 percentage points (6.40% vs. 5.79%) higher than that from other rotating groups, while the differences for labor force participation rates and employment-population ratios are 1.28% percentage points (66.46% vs. 65.18%) and 0.81% percentage points (62.22% vs. 61.41%), respectively.

To utilize the method present in the last section and estimate misclassification probabilities  $M_{U_4|U_4^*, U_3}$ , we need to obtain the joint distribution  $\Pr(U_{13}, U_4, U_3, U_1)$ . We follow the algorithm proposed by Madrian and Lefgren (2000) to match the CPS monthly files. Specifically, we first match the CPS sample based on household identifier, household replacement number, and personal identifier, then use the information on sex, age, and race to “certify” the crude matches. Nevertheless, such matched sample may not be representative of the cross-sectional sample due to sample attrition (Feng 2008). Therefore, we generate a matching weight to correct for sample attrition using the same procedure as in Feng and Hu (2013).<sup>17</sup> In order to have a large sample size, we pool all the matched samples over our study period to estimate misclassification probabilities.

<sup>17</sup>See detailed procedures on matching the monthly CPS data and correcting for sample attrition in Section A of the online Appendix.

Panel A of Table 1 shows the estimated misclassification probabilities for ongoing rotation groups.<sup>18</sup> When the current true status and the previously-reported status are the same, the probability of reporting that status is always the highest. For example, conditional on that the previously-reported status is employed, 98.8% of truly employed people would report being employed. We also find that even conditional on the current true status, the misclassification probabilities are still highly related to the previously-reported status. For example, conditional on that the previously-reported status is employed, 74.0% of truly unemployed individuals and 55.9% of truly not-in-labor-force individuals report being employed falsely.<sup>19</sup>

In addition, the results also suggest that when the current true status and the previously-reported status are different, the third status (different from both the current true status and the previously-reported status) becomes the least likely to be reported. This also confirms our re-ordering Assumption 5. For example, conditional on that the current true status is unemployed and the previously-reported status is employed, the probability of reporting being not-in-labor-force is only 6.3%, less than that of reporting being employed (74.0%) or unemployed (19.7%). When the current true status is not-in-labor-force and the previously-reported status is employed, the probability of reporting being unemployed is only 2.2%, less than that of reporting being employed (55.9%) or not-in-labor force (41.9%).

To make sure that our results on the persistence in misreporting behavior come from correlations between the reported statuses in two periods, we conduct a falsification test. The idea is that if the reported status is generated independently, we would not obtain the result of persistence as in Panel A of Table 1. After matching  $(U_{13}, U_4, U_3, U_1)$ , we fix the marginal distribution of  $U_3$  but randomly re-assign the reported status to generate  $(U_{13}, U_4, \tilde{U}_3, U_1)$ . This procedure makes the reported status  $U_4$  independent of the counterfactual previously-reported status  $\tilde{U}_3$ .<sup>20</sup> Panel B of Table 1 shows that there is no significant difference among the three misclassification matrices, implying that the current reported status is no longer affected by the previously-reported status conditional on the current true status.<sup>21</sup>

Once we have identified the misclassification probabilities for ongoing rotation groups, we can estimate the distribution of true LFS, which can then be used to back out the misclassification probabilities for rotation groups one and five. Panel C of Table 1 shows that the misclassification matrices of rotation group one and five are similar to those in earlier studies, in which the diagonal elements are always the largest in each column, i.e., people are more likely to report the truth. Note that this is no longer an imposed assumption as in previous studies. Similar to what was found in previous studies as in Abowd and Zellner (1985); Poterba and Summers (1986); Feng and Hu (2013), there exist considerable misclassification errors, especially for those whose true status is unemployed, with their probabilities of reporting being unemployed correctly to be only 61.0% and 57.1% for rotation groups one and five, respectively.

#### 4.2. Out-of-sample corrections and corrected labor force statistics

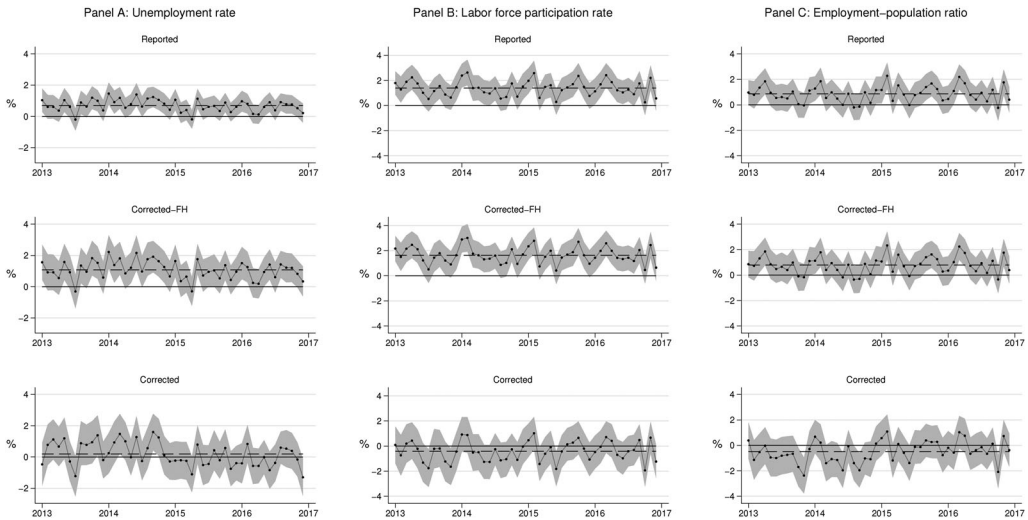
We next perform out-of-sample corrections to investigate whether our methodology eliminates RGB. For each period  $t$ , we first estimate misclassification probabilities for ongoing rotation

<sup>18</sup>We also control for observed heterogeneity and estimate  $\Pr(U_t|U_t^*, U_{t-1}, X)$ , and the patterns remain. See Section F of the online Appendix.

<sup>19</sup>Using a linked CPS-UI data set, Abraham et al. (2013) report that, among those workers who receive wages in the UI records ("true" status is employed), 6.4% do not report working in the CPS, while among those who do not receive wages in the UI records, 17.6% report working in the CPS. That is,  $\Pr(U = NE|U^* = E) = 6.4\%$  and  $\Pr(U = E|U^* = NE) = 17.6\%$ , where  $E$  and  $NE$  refer to "working" and "not working", respectively. In comparison, we integrate out previously-reported status in our misclassification probabilities and re-classify "unemployed" and "not-in-labor-force" as "not working", then calculate two corresponding probabilities,  $\Pr(U = NE|U^* = E) = 5.2\%$  and  $\Pr(U = E|U^* = NE) = 10.0\%$ , which are comparable with those in Abraham et al. (2013).

<sup>20</sup>Note that the partial relationship between  $U_4$  and  $U_3^*$ , if any, is not affected.

<sup>21</sup>A formal test based on bootstrap is shown in Table F7 of the online Appendix.

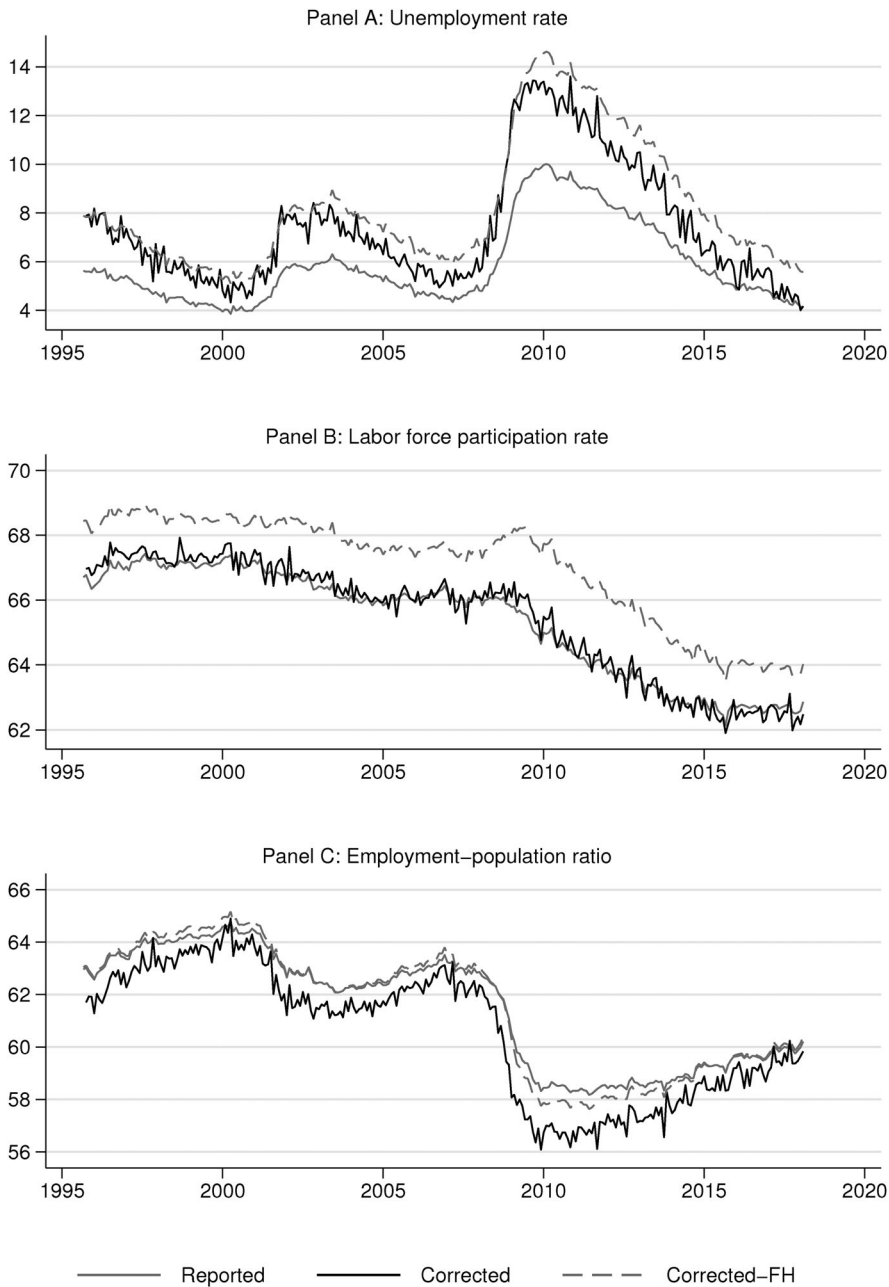


**Figure 2.** Out-of-sample corrections: differences of labor force statistics between rotation group one and other rotation groups. *Note:* “Reported” series are based on reported labor force statistics. “Corrected” series are calculated using the misclassification probabilities in this paper, while “Corrected-FH” series are calculated using the misclassification probabilities in Feng and Hu (2013). Each dot denotes the difference of labor force statistics between rotation group one and other rotation groups in a given month. The dashed horizontal lines are average values of the differences over the covered periods. The corresponding 95 percent confidence intervals are calculated from bootstrapped standard errors based on 500 repetitions.

groups, as well as for rotation groups one and five, using the CPS data from September 1995 up to period  $t - 1$ . Second, we use the estimated misclassification probabilities to correct the marginal distribution of LFS for period  $t$  and compare the labor force statistics from rotation group one and other rotation groups.

Panel A of Fig. 2 displays the results for unemployment rates from January 2013 to December 2016. The first row simply shows the existence of RGB in reported unemployment rates between rotation group one and other rotation groups. As we can see, almost all the differences lie above the horizontal zero-line. On average, the unemployment rate of rotation group one is 0.66 percentage points higher than that of other rotation groups for this time period. In the second row, we correct for misclassification errors using the method proposed in Feng and Hu (2013), where the misclassification probabilities are assumed to only depend on the current true status. Overall, the patterns regarding RGB are almost the same as the reported ones. In the third row, we report the correction results using our proposed method in this paper. On average, RGB almost disappears completely, as the average difference between rotation group one and other rotation groups is only 0.12 percentage points, compared to 0.66 percentage points in the first row and 1.02 percentage points in the second row. Similar results also appear in the out-of-sample corrections for labor force participation rates and employment-population ratios, as shown in Panels B and C, respectively.

Although misclassification errors in the CPS have been realized and studied before, these studies have ignored the issue of RGB and not incorporated persistence in the reporting process, such as Abowd and Zellner (1985); Poterba and Summers (1986); Feng and Hu (2013). Figure 3 compares our corrected labor force statistics with the reported series, as well as the corrected series based on Feng and Hu (2013). In terms of levels, Panel A shows that the “Reported” unemployment rates are lower than our “Corrected” series on average, which shows that the official unemployment rates systematically underestimate true levels of unemployment, consistent with the main findings of Feng and Hu (2013). However, our “Corrected” series, with an average of 7.43%, is lower than the “Corrected-FH” series in levels (8.26%). Therefore, ignoring persistence



**Figure 3.** Reported and corrected labor force statistics.

*Note:* “Reported” series are reported labor force statistics. “Corrected” series are calculated using the misclassification probabilities in this paper, while “Corrected-FH” series are calculated using the misclassification probabilities in Feng and Hu (2013). All numbers are seasonally-adjusted using the Census’s X-12-ARIMA algorithm.

in reporting errors leads to an overestimation of unemployment rates. In addition, Panel B shows that our “Corrected” labor force participation rates nearly coincide with the “Reported” series, which is substantially lower than the “Corrected-FH” series. Similarly, in Panel C, our “Corrected” employment–population ratios are also lower than both the “Reported” and “Corrected-FH” series. Nevertheless, a reassuring pattern is that all these series track very

**Table 2.** Average rotation group bias in labor force statistics over 1995/09–2018/02.

	Reported				Corrected
	All RGs	RG one	Other RGs	RGB	
Unemployment rate	5.9 (0.01)	6.4 (0.02)	5.8 (0.01)	0.6 (0.02)	7.4 (0.19)
Labor force participation rate	65.3 (0.01)	66.4 (0.03)	65.2 (0.01)	1.3 (0.03)	65.4 (0.09)
Employment-population ratio	61.5 (0.01)	62.2 (0.03)	61.4 (0.01)	0.8 (0.03)	60.6 (0.11)

Note: In parentheses are bootstrapped standard errors based on 500 repetitions.

**Table 3.** Misclassification probabilities by type of respondent,  $\Pr(U_4 = i | U_4^* = j, U_3)$ .

	$U_3 = 1$			$U_3 = 2$			$U_3 = 3$		
	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
Panel A: Self-reported respondents									
$i = 1$	99.2 (0.02)	78.3 (0.86)	57.5 (0.75)	49.5 (0.85)	4.7 (0.48)	4.2 (0.47)	45.2 (0.57)	2.6 (0.55)	0.5 (0.01)
$i = 2$	0.4 (0.02)	18.4 (0.75)	2.7 (0.46)	44.0 (0.88)	88.4 (1.26)	43.2 (2.20)	7.3 (0.38)	43.8 (1.67)	0.3 (0.03)
$i = 3$	0.4 (0.01)	3.4 (0.58)	39.8 (0.94)	6.6 (0.38)	6.9 (1.16)	52.6 (2.15)	47.5 (0.51)	53.6 (1.56)	99.2 (0.04)
Panel B: Proxy-reported respondents									
$i = 1$	98.6 (0.03)	73.2 (0.97)	57.9 (0.61)	46.6 (0.74)	4.2 (0.73)	7.7 (0.47)	37.7 (0.38)	5.2 (0.49)	1.0 (0.02)
$i = 2$	0.5 (0.02)	19.0 (0.86)	1.9 (0.50)	42.2 (0.77)	95.8 (0.73)	41.6 (2.24)	6.1 (0.24)	39.0 (1.52)	0.3 (0.06)
$i = 3$	0.9 (0.02)	7.8 (0.97)	40.2 (0.91)	11.2 (0.46)	0.0 (0.00)	50.7 (2.02)	56.1 (0.38)	55.8 (1.50)	98.7 (0.07)

Note: Panel A is estimated using the self-reported samples in all four periods in the matched data  $(U_{13}, U_4, U_3, U_1)$ , while Panel B is from proxy-reported samples. In parentheses are bootstrapped standard errors based on 500 repetitions.

similarly, implying that labor market developments over time appear to be robust to misclassification errors.

## 5. Discussions

### 5.1. Comparison with Ahn and Hamilton (2022)

Recently, Ahn and Hamilton (2022) also correct for RGB and other inconsistencies in the CPS data, including sample attrition, misclassification, and unemployment duration. In terms of RGB, they argue that it is related to the number of times interviewed and interpret this as differences in interview technology among rotation groups. Ahn and Hamilton's framework is essentially considering the following counterfactual question: how would labor force statistics have changed if a rotation group had instead been interviewed using the interview technology of another rotation group. They use rotation group one as a benchmark for labor force statistics of other rotation groups, which is supported by several aspects of empirical evidence, including survey disengagement, stigma, and so on. In comparison, our model does not impose any assumption regarding which rotation group would report more accurate labor force statistics, nor does it have any *a priori* implications on the existence or direction of RGB.

Regarding misclassification errors, Ahn and Hamilton (2022) utilize the information of reported unemployment duration. In particular, they re-classify *UNU* transition as *UUU* if the final *U* reports a duration of job search greater than 4 weeks, as they find that the job-finding probabilities reported by people who make *NU* and *UU* transitions with unemployment duration



above 4 weeks are essentially the same.<sup>22</sup> However, Ahn and Hamilton only focus on the misclassified  $N$ , ignoring that other labor force statuses are also subject to potential misclassification errors. For instance, a validation study has shown 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 (Abraham et al. 2013). Instead, our approach considers potential misclassification across all three labor force statuses jointly.

Although the two papers utilize very different frameworks, the corrected unemployment rates are similar, with an average of 7.9% over 2001–2018 in this paper and 8.2% in Ahn and Hamilton (2022). We also find the reported unemployment rates from rotation group one closest to the corrected unemployment rates, which is consistent with Ahn and Hamilton's choosing rotation group one as a benchmark in correcting for RGB.

## 5.2. Potential mechanisms

In this subsection, we discuss theories in the “panel conditioning” literature that may shed light on why RGB arises in the CPS. One possible explanation is the “practice effect”, that is, respondents may gain a better understanding of survey questions as subsequent interviews continue. According to this conjecture, later rotation groups should be closer to the truth than rotation group one. However, comparisons from Table 2 show that the reported unemployment rate from later rotation groups deviates more from the corrected unemployment rate. This suggests that unemployment tends to be under-reported systematically in all rotation groups, and repeated interviewing leads to more inaccuracy progressively. This pattern is robust for different sub-periods, demographic groups, and regions.<sup>23</sup> Therefore, these empirical results cast doubt on the hypothesis that RGB arises through the “practice effect”.

On the other hand, two other hypotheses are potentially consistent with the pattern of progressively more misclassification errors, both related to survey manipulation by respondents. The first is that respondents are more likely to avoid stigmatized answers such as being unemployed (Halpern-Manners and Warren 2012). As Ahn and Hamilton (2022) note, self-reported respondents may care more about stigmatization than the proxy-reported. Based on our estimates in Table 3, conditional on that the previously-reported status is unemployed, the probability of truly unemployed people's reporting being not-in-labor-force is higher for the self-reported (6.9%) than the proxy-reported (0.0%). Ahn and Hamilton (2022) also find that unemployment falls more quickly across rotation groups among self-reported respondents than the proxy-reported.

The second hypothesis, as suggested in Halpern-Manners and Warren (2012), is that respondents may become disengaged from repeated interviews and strategically choose answers that may lead to less interview time. Using records on household's cumulative interview time in the CPS, we find that an increase in the number of unemployed people in a household is associated with an increase in interview time by 12%, as shown in Table 4. This suggests that there is an incentive for respondents to not report being unemployed if they want to reduce interview time. Similarly, our results show that conditional on household size, total interview time is negatively related to the number of disabled people in a household. This is consistent with the findings in Ahn and Hamilton (2022) that people in later rotation groups are more likely to report being not-in-labor-force for the reason of disability, and disabled people in later rotation groups are more likely to return to the labor force in the future than those in rotation group one. In the meantime, the number of multiple job holder is positively related to interview time, consistent

<sup>22</sup> $E$ ,  $U$  and  $N$  stand for employment, unemployment and not-in-labor-force, respectively. This procedure differs from that in Rothstein (2011); Farber and Valletta (2015); Elsby, Hobijn, and Şahin (2015) among others, which re-classify all  $UNU$  transition as  $UUU$ .

<sup>23</sup>See Section F of the online Appendix.



**Table 4.** The impact of unemployment on household's cumulative interview time.

	(1)	(2)	(3)	(4)
	Interview time		Log (interview time)	
# of unemployed in household	144.1*** (1.352)	83.10*** (1.675)	0.222*** (0.002)	0.119*** (0.002)
# of disabled in household	-87.37*** (1.155)	-49.50*** (2.535)	-0.169*** (0.002)	-0.0865*** (0.003)
# of multi-jobs holder in household	176.9*** (1.170)	126.0*** (1.786)	0.279*** (0.002)	0.166*** (0.002)
Rotation group dummies (RG1 as reference group)				
RG2	-446.7*** (1.388)	-458.8*** (1.276)	-0.707*** (0.002)	-0.728*** (0.002)
RG3	-478.6*** (1.388)	-493.9*** (1.291)	-0.777*** (0.002)	-0.804*** (0.002)
RG4	-398.7*** (1.390)	-417.1*** (1.309)	-0.631*** (0.002)	-0.664*** (0.002)
RG5	-245.7*** (1.397)	-262.5*** (1.716)	-0.316*** (0.002)	-0.349*** (0.002)
RG6	-483.2*** (1.393)	-506.2*** (1.752)	-0.783*** (0.002)	-0.832*** (0.002)
RG7	-497.2*** (1.392)	-524.1*** (1.787)	-0.817*** (0.002)	-0.875*** (0.002)
RG8	-413.1*** (1.392)	-443.0*** (1.817)	-0.661*** (0.002)	-0.726*** (0.002)
Constant	841.4*** (2.353)	870.9*** (2.830)	6.498*** (0.003)	6.576*** (0.004)
Household size dummies	Yes	Yes	Yes	Yes
Household fixed effects	No	Yes	No	Yes
Observations	1,858,167	1,858,167	1,858,167	1,858,167
R-squared	0.285	0.283	0.331	0.359
Number of household	386,765	386,765	386,765	386,765

Note: We only use the CPS data from September 1995 to December 1998, as there is no record on household's cumulative interview time since 1999. \*\*\*, \*\*, \* signify the difference is statistically significant at the 1%, 5%, 10% level, respectively.

with the findings in Hirsch and Winters (2016) that people in rotation group one are more likely to report holding multiple jobs. All of these are consistent with the hypothesis of avoiding lengthy interviews. However, additional research, preferably controlled experiments, is necessary to provide more credible causal evidence.

## 6. Conclusion

In this paper, we develop a general statistical framework to study the issue of rotation group bias. In particular, we allow the distribution of misclassification errors to depend not only on current true status, but also on previously-reported status. Under certain conditional independence assumptions, we identify the distributions of misclassification errors for different rotation groups, with which we correct for the misclassification errors in labor force status and calculate the corrected labor force statistics.

Using the CPS data, we find strong evidence for our key assumption that both the current true status and the previously-reported status matter. In other words, what respondents reported in the previous interview does affect the reported value in the current interview. The out-of-sample corrections show that our proposed method does eliminate rotation group bias. We also provide new estimates of the U.S. unemployment rates, which are much higher than the official series, but lower than the previous estimates that ignored persistence in misclassification.

Although this paper specifically focuses on the issue of rotation group bias in the CPS, many surveys have similar rotating structures as the CPS and our method can be applied directly there with some modification. More generally, researchers who take misclassification errors seriously can extend our proposed model to many empirical studies using panel data.

## Funding

Feng acknowledges support from the National Natural Science Foundation of China (Project Number: 72073052) and the National Science Fund for Distinguished Young Scholars of China (Project Number: 71425005).

## References

- Abowd, J. M., Zellner, A. (1985). Estimating gross labor-force flows. *Journal of Business & Economic Statistics* 3(3): 254–283. doi:10.2307/1391596
- Abraham, K. G., Haltiwanger, J., Sandusky, K., Spletzer, J. R. (2013). Exploring differences in employment between household and establishment data. *Journal of Labor Economics* 31(S1):S129–S172. doi:10.1086/669062
- Ahn, H. J., Hamilton, J. D. (2022). Measuring labor-force participation and the incidence and duration of unemployment. *Review of Economic Dynamics* 44:1–32. doi:10.1016/j.red.2021.04.005
- Bailar, B. A. (1975). The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association* 70(349):23–30. doi:10.1080/01621459.1975.10480255
- Bailar, B. A. (1989). Information needs, surveys, and measurement errors. In: Kasprzyk, D., Duncan, G. J., Kalton, G., Singh, M.P., eds., *Panel Surveys*, New York: John Wiley & Sons, pp. 1–24.
- Bassi, F., Trivellato, U. (2009). A latent class approach for estimating gross flows in the presence of correlated classification errors. In: Peter, L., ed., *Methodology of Longitudinal Surveys*, Chichester, UK: John Wiley & Sons, Ltd, pp. 367–380.
- 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.
- Bollinger, C. R., David, M. H. (2005). I didn't tell, and I won't tell: Dynamic response error in the SIPP. *Journal of Applied Econometrics* 20(4):563–569. doi:10.1002/jae.794
- Bound, J., Brown, C., Mathiowetz, N. (2001). Measurement error in survey data. In: Heckman, J. J., Leamer, E., eds., *Handbook of Econometrics, Volume 5 of Handbooks in Economics*, Amsterdam: Elsevier Science, Chapter 59, pp. 3705–3843.
- Chua, T. C., Fuller, W. A. (1987). A model for multinomial response error applied to labor flows. *Journal of the American Statistical Association* 82(397):46–51. doi:10.1080/01621459.1987.10478389
- Elsby, M. W., Hobbijn, B., Şahin, A. (2015). On the importance of the participation margin for labor market fluctuations. *Journal of Monetary Economics* 72:64–82. doi:10.1016/j.jmoneco.2015.01.004
- Erkens, G. (2012). Changes in panel bias in the US Current Population Survey and its effects on labor force estimates. In: *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 4220–4232. [http://www.asasrms.org/Proceedings/y2012/Files/304857\\_73851.pdf](http://www.asasrms.org/Proceedings/y2012/Files/304857_73851.pdf)
- 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. doi:10.3368/jhr.50.4.873
- 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. doi:10.3233/JEM-2008-0306
- Feng, S., Hu, Y. (2013). Misclassification errors and the underestimation of the US unemployment rate. *The American Economic Review* 103(2):1054–1070. doi:10.1257/aer.103.2.1054
- Ghangurde, P. D. (1982). Rotation group bias in the LFS estimates. *Survey Methodology* 8:86–101.
- Halpern-Manners, A., Warren, J. R. (2012). Panel conditioning in the Current Population Survey: Implications for labor force statistics. *Demography* 49(4):1499–1519. doi:10.1007/s13524-012-0124-x
- Hansen, M. H., Hurwitz, W. N., Nisselson, H., Steinberg, J. (1955). The redesign of the Census Current Population Survey. *Journal of the American Statistical Association* 50(271):701–719. doi:10.1080/01621459.1955.10501962
- Hirsch, B. T., Winters, J. V. (2016). Rotation group bias in measures of multiple job holding. *Economics Letters* 147:160–163. doi:10.1016/j.econlet.2016.08.039
- 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. doi:10.1016/j.jeconom.2007.12.001
- 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. doi:10.1016/j.jeconom.2017.06.002
- Hu, Y., Shum, M. (2012). Nonparametric identification of dynamic models with unobserved state variables. *Journal of Econometrics* 171(1):32–44. doi:10.1016/j.jeconom.2012.05.023
- Irvine, J. M. (1984). Rotation group bias and the Methods Test Panel. In: *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 714–719. [http://www.asasrms.org/Proceedings/papers/1984\\_135.pdf](http://www.asasrms.org/Proceedings/papers/1984_135.pdf)

- Krueger, A. B., Mas, A., Niu, X. (2017). The evolution of rotation group bias: Will the real unemployment rate please stand up? *The Review of Economics and Statistics* 99(2):258–264. doi:10.1162/REST\_a\_00630
- 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. doi:10.3233/JEM-2000-0165
- Magnac, T., Visser, M. (1999). Transition models with measurement errors. *Review of Economics and Statistics* 81(3):466–474. doi:10.1162/003465399558229
- McCarthy, P. J. (1978). Some sources of error in labor force estimates from the Current Population Survey. Background Paper No. 15, Washington, DC: U.S. Government Printing Office (National Commission on Employment and Unemployment Statistics).
- Poterba, J. M., Summers, L. H. (1986). Reporting errors and labor market dynamics. *Econometrica* 54(6): 1319–1338. doi:10.2307/1914301
- Poterba, J. M., Summers, L. H. (1995). Unemployment benefits and labor market transitions: A multinomial logit model with errors in classification. *The Review of Economics and Statistics* 77(2):207–216. doi:10.2307/2109860
- Rothstein, J. (2011). Unemployment insurance and job search in the Great Recession. *Brookings Papers on Economic Activity* 2011(2):143–196. doi:10.1353/eca.2011.0018
- Shibata, I. (2019). Are labor market indicators telling the truth? Role of measurement error in the U.S. Current Population Survey. International Monetary Fund Working Paper 19/40. <https://www.imf.org/en/Publications/WP/Issues/2019/02/26/Are-Labor-Market-Indicators-Telling-the-Truth-Role-of-Measurement-Error-in-the-U-S-46597>
- Shockey, J. W. (1988). Adjusting for response error in panel surveys: A latent class approach. *Sociological Methods & Research* 17(1):65–92. doi:10.1177/0049124188017001004
- Silverstone, B., Bell, W. (2011). Gross labour market flows in New Zealand: Some questions and answers. Department of Economics Working Paper 15/11. Hamilton, New Zealand: University of Waikato. <https://researchcommons.waikato.ac.nz/handle/10289/6700>
- 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. doi:10.1080/01621459.1996.10476966
- Solon, G. (1986). Effects of rotation group bias on estimation of unemployment. *Journal of Business & Economic Statistics* 4(1):105–109.
- Sturgis, P., Allum, N., Brunton-Smith, I. (2009). Attitudes over time: The psychology of panel conditioning. In: Lynn, P., ed., *Methodology of Longitudinal Surveys*, New York: Wiley Online Library, pp. 113–126.
- U.S. Bureau of the Census. (1978). The Current Population Survey: Design and Methodology. Technical Paper 40, U.S. Bureau of the Census.
- Van den Brakel, J. A., Krieg, S. (2015). Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design. *Survey Methodology* 41(2):267–296.
- Warren, J. R., Halpern-Manners, A. (2012). Panel conditioning in longitudinal social science surveys. *Sociological Methods & Research* 41(4):491–534. doi:10.1177/0049124112460374
- Waterton, J., Lievesley, D. (1989). Evidence of conditioning effects in the Social Attitudes Panel. In: Kasprzyk, D., Duncan, G. J., Kalton, G., Singh, M. P., eds., *Panel Surveys*, New York: John Wiley & Sons, pp. 319–339.
- Williams, W. H., Mallows, C. L. (1970). Systematic biases in panel surveys due to differential nonresponse. *Journal of the American Statistical Association* 65(331):1338–1349. doi:10.1080/01621459.1970.10481169
- Zwane, A. P., Zinman, J., Van Dusen, E., Pariente, W., Null, C., Miguel, E., Kremer, M., Karlan, D. S., Hornbeck, R., Giné, X., Duflo, E., Devoto, F., Crepon, B., Banerjee, A. (2011). Being surveyed can change later behavior and related parameter estimates. *Proceedings of the National Academy of Sciences of the United States of America* 108(5):1821–1826. doi:10.1073/pnas.1000776108