



# Masking Real Unemployment: The Overall and Racial Impact of Survey Non-Response on Measured Labor Market Outcomes

Yixia Cai and Dean Baker\*

Working Paper No. 150

February 21<sup>st</sup>, 2021

## ABSTRACT

A large and growing percentage of households are missed in the monthly Current Population Survey (CPS). For the survey as a whole, the rate of nonresponse is roughly 13 percent. This is higher for Blacks, with the share for young Black men being about 30 percent. The BLS's current methodology effectively assumes that, with adjustment for various characteristics, people who are not included in a follow-up survey may not differ systematically from those who are

---

\* Yixia Cai is an economist at the Center for Economic and Policy Research (CEPR). Dean Baker is co-founder and senior economist at CEPR. We thank Heidi Shierholz, John Schmitt for helpful comments on an earlier version of the paper. The support from Institute for New Economic Thinking is greatly appreciated. All errors are our own. Direct correspondence to Yixia Cai; [cai@cepr.net](mailto:cai@cepr.net). Data and codes of the paper are available upon request.

included. The present paper, however, provides evidence that this may not be the case. With the rotation panel structure of the CPS data from 2003 to 2019, we investigate bias from nonresponse in CPS and its association with one's prior labor market status, paying particular attention to how the relationship differs by race, ethnicity, and gender. Our analysis suggests that people are considerably more likely to be missing in a subsequent observation if they are unemployed or not in the labor force in the prior observation. We also estimate what the real labor market outcomes might have been when adjusting for nonresponse and undercoverage. Findings indicate that the current methodology may underestimate the national unemployment and labor force participation rates by about 0.7 and 0.5 percentage points, respectively. The gap between observed and adjusted unemployment rates tends to grow beginning in 2015. The unemployment rate is more understated for Blacks than for whites, particularly with a gap of about 3.3 percentage points for young Black men (age 16 to 34). The unemployment rate for Black women is understated by around 2.4 percentage points.

**<https://doi.org/10.36687/inetwp150>**

**JEL codes:** J01, J64, J70

**Key words:** racial discrimination, unemployment, Black men, racial employment gaps

## Executive Summary

A large and growing percentage of households are missed in the monthly Current Population Survey (CPS). For the survey as a whole, the rate of nonresponse is roughly 13 percent. This is higher for Blacks, with the share for young Black men being over 30 percent. The Bureau of Labor Statistics' (BLS) current methodology effectively assumes the people who are not included in the survey do not differ systematically from those who are included when controlling for various characteristics, such as race, age, and gender. The present paper, however, provides evidence that this may not be the case.

Using the rotation panel structure of the CPS 2003 to 2019, we investigate bias from nonresponse in CPS and its association with one's prior labor market status. We also look at how the impact differs by race, ethnicity, and gender. The analysis suggests that people are considerably more likely to be missing in a subsequent observation, if they are either unemployed or not in the labor force in the prior observation. Not surprisingly, people are also far more likely to remain unemployed or out of the labor force in two consecutive months. We impute labor market status for the missing observations based on their labor force state in previous and subsequent observations (if any), taking into account other characteristics. We then apply this estimate to all the non-responders, since we only have prior observations for around one-third of the non-responders. The other two-thirds were either never in the survey (there was no response for any observation) or were missing for multiple observations (e.g. they responded the first time, but then were missed the second, third, and fourth time).

We used two different methodologies to make adjustments. First, we used a single imputation, where we imputed labor market status for missing observations, based exclusively on their prior observation. We then used an iterative process that estimated the labor market status for missing observations based on all the observations for which we had responses. (This procedure is described in the text.)

The multiple imputation adjustment leads to:

- An overall unemployment rate, for the years 2003 to 2019, that is 0.7 percentage points higher than with the current methodology.
- An overall employment-to-population rate (EPOP) that is the same as with the current methodology.
- An unemployment for Blacks that is 2.6 percentage points higher than with the current methodology.
- An EPOP for Blacks that is 1.6 percentage points lower than with the current methodology.
- An unemployment rate for Hispanics that is 1.5 percentage points higher than with the current methodology.
- An EPOP for Hispanics that is 0.3 percentage points lower than with the current methodology.
- An unemployment rate for young Black men between the ages of 16 to 24 that is 3.6 percentage points higher than with the current methodology.
- An employment rate for this age group that is 1.7 percentage points lower than with the current methodology.

- An unemployment rate for young Black men between the ages of 25 to 34 that is 3.0 percentage points higher than the with the current methodology.
- An employment rate for this age group that is 2.0 percentage points lower than with the current methodology.
- An unemployment rate for Black women that is 2.4 percentage points higher than with the current methodology.
- An employment rate for Black women that is 1.4 percentage points lower than with the current methodology.

This analysis supports prior research that suggests the current methodology for treating missing observations may lead to an understatement of the true unemployment rate, and the bias is especially salient for those who are most disadvantaged in the labor market. As coverage rates in the CPS fall through time, this understatement is likely to grow larger over time.

## Introduction

There is renewed interest in determining the real labor market hardships faced by many in the United States despite its strong economy. The government's surveys play a central role in providing information on this topic. The Census Bureau's Current Population Survey (CPS) is the underlying source of many official labor market statistics, as well as income and poverty measures, and health insurance coverage. However, recent research has raised questions about its accuracy in certain areas.

Unit nonresponse in major surveys has drawn increasing attention in the past decade, including in the CPS (Ahn and Hamilton 2020; Bailer 1975; Meyer, Mok, and Sullivan 2015; Schmitt and Baker 2005, 2006). Prior literature finds evidence of systemic bias in the monthly CPS data, with notable differences between the months in which respondents are in the survey (Krueger, Mas, and Niu 2017).

This paper examines the potential bias that could result from missing housing units or observations, which results in undercoverage<sup>1</sup> in the CPS. Undercoverage was estimated to be around 11 percent in 2019, although the exact rate varies monthly and by demographic characteristics (U.S. Census Bureau 2019). If undercoverage biases estimates of labor force statistics, it could have substantial implications for our understanding of labor-market conditions both at a time point and trends through time. This analysis investigates both the extent to which nonresponse bias may impact the actual (un)employment statistics in general, and the extent to which it affects estimates for Black workers.

There are three primary issues we address in the present paper. First, in discussing any bias in employment statistics, we ask whether and to what extent someone's unemployment status could affect their (non)response in the next survey period. Second, we assess whether the effects of survey nonresponse bias vary across different ages for Black, Hispanic, and white workers, paying particular attention to younger men of racial and ethnic minorities. Lastly, we provide a rough estimation of what the national unemployment rate, employment-to-population ratio, and labor-force participation rate would look like after considering nonresponse and undercoverage bias.

Recent articles argue that unemployment after the onset of the pandemic would have been much higher than reported (Ahn and Hamilton 2020; Rothbaum and Bee 2020; Ward and Edwards 2020). Undoubtedly, the sudden economic shutdown and the resulting labor market consequences, as well as drastic racial disparity of responses, may pose tremendous challenges to the representativeness of the recent monthly data. Therefore, a more fundamental issue regarding

---

<sup>1</sup> In the subsequent sections, we use the term "coverage ratio," which is defined as the estimate from the sample of the number of people for a specific characteristic divided by the same estimate from the decennial census. For details, see <https://www.census.gov/programs-surveys/cps/technical-documentation/methodology/coverage-ratios.html>

the underestimation of labor force statistics inevitably has to deal with monthly nonresponses. Indeed, this underreporting or nonresponse issue in the survey may have existed for a long time, but happened to be catalyzed by the disrupted labor market situation introduced by the pandemic. Following work by Bailar (1975), Krueger and colleagues (2014), for example, document that the unemployment rate rises more sharply for participants surveyed in earlier months compared to ones surveyed in later months, suggesting a systematic bias in unemployment statistics depending on the months in which someone is in the sample. The authors further signal such bias is very likely attributed to the nonresponse rate in the survey.

In theory, if survey nonresponse happens to be random, then any estimate of labor force statistics that simply excludes non-respondents will still tend to be unbiased. Yet, it may not be straightforward to validate whether nonresponse across the population differs from that among survey respondents, partly because information is lacking about the characteristics of these missing observations and because no universal consensus exists of what the true labor force statistics would be. Researchers have utilized different approaches to measure nonresponse bias in the CPS by using matching data from the Decennial Census (Schmitt and Baker 2006) or taking advantage of the longitudinal design of CPS data (Ahn and Hamilton 2020; Krueger et al. 2017). Earlier works, for example, documented that the CPS may lose a share of nonworking adults over time (Schmitt and Baker 2006). More recently, Ahn and Hamilton (2020) utilized the CPS's longitudinal design to investigate interview skipping and concluded that CPS may likely understate unemployment rate and labor force participation rate by about 2 percentage points.

In the present paper, we investigate whether nonresponse bias is linked with prior employment status and decompose the overall bias into subgroups. We find that unemployment in the current period results in a statistically significant increase in the probability of missing observations in the next month's survey, and Black workers are more likely to have missing observations, as compared to their white counterparts. Younger workers also tended to be less likely to respond to follow-up surveys. With imputation and adjustment for the potential undercoverage issue and nonresponse bias, we find that the average unemployment rate is underestimated by 0.7 percentage points and that the labor force participation rate would have been 0.5 points higher over the period from 2003 to 2019. The differences tend to be far larger for Black workers, particularly young Black men.

## **Data and Method**

Our micro data come from the monthly CPS survey spanning 2003 to 2019, which include civilian noninstitutional population<sup>2</sup>. One signature design of the survey is its rotation group panel feature. In theory, a respondent appears in the sample for a total of eight months. Respondents are initially interviewed for four consecutive months, then leave for eight months, and are then re-interviewed for another four months. Yet, a non-negligible portion of respondents

---

<sup>2</sup> We access the data through CPS-IPUMS database (Flood et al. 2020).

appear to be missing for one or more months in the survey. Some leave the sample but return, and others drop out of the survey completely after their first or second appearance in the survey. Though various reasons exist for a surveyed respondent to not have a follow-up observation, prior studies on similar topics indicate good reasons to believe that one's prior labor market status could potentially impact their dropping out of the survey (Ahn and Hamilton 2020; Krueger et al. 2017; Schmitt and Baker 2006). To minimize the impact of recent unprecedented job losses during the pandemic on our estimates, we exclude data from January 2020 and onward.

Taking advantage of the survey's panel design, we estimate the probability of missing the next month based on one's employment status in the present month. To determine whether unemployment is associated with next month's nonresponse in the survey, we first estimate the following specification:

$$M_{it+1} = \alpha + \delta_1 UNE_{it} + \delta_2 NLF_{it} + \beta_1 X_{it} + \beta_2 Y_{it} + \theta_t + \mu_i + \varepsilon_i \quad (1)$$

where  $M_{it+1}$  is a dichotomous variable with 1 indicating respondent  $i$  is missing at time  $t+1$ .  $UNE_{it}$  and  $NLF_{it}$  are two binary indicators of whether person  $i$  is unemployed or not in the labor force at time  $t$ . The coefficients of interest in this specification is  $\delta_1$  and  $\delta_2$ . As mentioned above, if non-respondents do not differ from the surveyed respondents in terms of employment status, then  $\delta_1$  and  $\delta_2$  should be non-significant.  $X_i$  is a set of individual characteristics, including race/ethnicity, age, gender, and education level<sup>3</sup>.  $Y_i$  is a set of dummy variables indicating types of occupation<sup>4</sup> for the employed.  $\theta_t$  denotes a set of year fixed effects.  $\mu_i$  indicates state of residence. Further, we are interested in whether nonresponse effects of employment status vary by certain subgroups. Toward this end, we estimate the following specifications:

$$M_{it+1} = \alpha + \delta_1 UNE_{it} + \beta_1 Age_{it} + \delta_2 (UNE * Age)_{it} + \varepsilon_i \quad (2)$$

where we regress the nonresponse indicator at time  $t+1$  on the unemployment status at time  $t$ , participants' ages, and the interaction between the two.  $\delta_2$  tests the different effects of unemployment on dropping out of the survey across age groups. We run models separately for each of the demographic groups (Black men, Black women, Hispanic men, Hispanic women, white men, and white women):

Because the missing observations at time  $t+1$  and its potential labor-market state are the main focuses in the paper, we proceed carefully with estimating their labor force status. Two approaches are used to impute the labor-force state of the missing observations: single imputation (SI) and multiple imputation (MI). First, we assign one of three labor market states – employed, unemployed, or not in the labor force – to the missing observations at time  $t+1$  based on their labor-force status at time  $t$ . Under this scenario, it is assumed that the transitions for the

---

<sup>3</sup> Race/ethnicity consists of four categories representing worker  $i$ 's race or ethnicity. Age is a categorical variable denoting worker  $i$ 's age category, namely 16-24, 25-34, 35-44, and 45-54, with 35-44 being the reference group. We exclude respondents (age 55 or over) in all models because nonresponse does not seem to be a large issue for this group given its coverage ratio of about 95 percent. Education includes three categories: lower education (high school diploma or less), some college, and college or higher education. Some of these coefficients are of interest as well, as they may signal the extent to which each subgroup is more or less likely to become nonresponses in the next period.

<sup>4</sup> We construct occupation types using the four-digit occupation variable `occ2010` provided by IPUMS.

missing observations are the same as the transitions for people with the same labor market status at time  $t$ , for whom we have observations.

It is possible that this imputation is not accurate because other factors may affect the non-response (Dziura et al. 2013; Rubin 2004). For example, someone experiencing a labor-market-status change from time  $t$  to  $t+1$  might also impact the response rate at  $t+1$ . A respondent transitioning into employment at  $t+1$  may miss the present survey interview as he or she may be quite involved with starting the new job, before returning to the survey next month. Additionally, someone with a transition from employment to unemployment at time  $t+1$  may be less or more likely to answer the survey at time  $t+1$ . Insofar as this is the case, our single imputation method could overstate the extent of unemployment and non-employment for missing observations.

Second, we use MI to handle the missing observations. This method helps us to understand an individual's labor-market status from multiple observations that may jointly predict his or her status in the missing observation. This method also allows us to understand how an individual's labor-market status from multiple observations may jointly predict his or her status in the missing observation. We conduct the MI through chained equations (MICE). This is an iterative process; in each round, four multinomial logistic models are separately used to predict each of the four observations, which takes the following form<sup>5</sup>:

$$P(Y_1 | X, Y_{\sim 1}, \varphi_1) \dots \dots P(Y_j | X, Y_{\sim j}, \varphi_j) \quad (3)$$

where  $Y = (Y_1, Y_2, \dots Y_j)$  is  $j$ th labor market outcome variables to be imputed<sup>6</sup>. Each  $Y$  is an unordered categorical labor market variable; either employed, unemployed, or not in the labor force. Let  $X$  be a vector of covariates, including someone's race/ethnicity, age, gender, and education. The predictors in each model include a respondent's labor market status in previous and subsequent observations (if any), along with the aforementioned covariates. We impute five datasets, and the estimates averaged across these datasets are nearly consistent with the non-imputed data (see appendix [panel A in Table A1])<sup>7</sup>. Panel B in Table A1 reports three labor-market outcomes for the missing observations at  $t+1$  under two approaches.

Although correcting group miscategorization is beyond the scope of this paper, it has been suggested that a number of respondents were miscategorized as not being in the labor force at period  $t$ , even if they reported being unemployed at  $t+1$  and stated that they had been looking for a job for over 4 weeks (Ahn and Hamilton 2020). In our sample, we find a disproportionate misclassification among non-Hispanic Black and Hispanic respondents who were counted as not in the labor force at  $t$  but as unemployed at  $t+1$ , while simultaneously reporting at time  $t+1$  that they were continuously unemployed for 5 weeks or more. This only considers the issue between  $t$

---

<sup>5</sup> It started with the least amount of missing values as the first outcome based on all observed data in the model and filled in imputed values for that variable. It then estimated the imputation model for the outcome with the second least amount of missing values, using both the observed data and the imputed values of the first variable. This process continued for all four labor market outcomes with missing values. Ten iterations were performed.

<sup>6</sup> As each respondent is supposed to have four continuous time points,  $j$  is four in this case.

<sup>7</sup> We understand that increasing the number of imputed completed datasets would be preferable because doing so would reduce sampling variability from the imputation procedure. However, given that we only considered respondents' missing observations at  $t+1$  and that the percentage of such missingness is less than 4 percent in the analytic sample, we only imputed five datasets. We will consider imputing more datasets to increase the reliability and including more factors in the imputation model.



and  $t+1$ , but the misclassification may occur at  $t+2$  if respondents reported having been unemployed for longer than 8 or 9 weeks.

Next, we use the preferred MI approach and its estimates for the subsequent discussion. A table reporting the new adjusted labor-market outcomes under the SI approach is also included in the appendix. With an assumption that those who were not covered are similar to those who had filled out prior observations but were missing in the next observations in ways other than race/ethnicity, gender and age, we calculate a new set of unemployment<sup>8</sup>, employment, and labor force participation rates. To calculate monthly adjusted (un)employment rates for each race-gender-age cell, we first calculate monthly unemployment rates of the surveyed by multiplying each month's race-gender-age shares of unemployed people in the labor force by monthly survey coverage ratio<sup>9</sup> for each race-gender-age group. We then calculate monthly unemployment rates of the non-surveyed by multiplying monthly imputed race-gender-age unemployment rate of the missing observations by monthly fraction of population not covered for each race-gender-age group. The sum of the two is applied for each race-gender-age as follows:

$$Z_{yt,g} = CR_{yt,g} * LMS_{obs,yt,g} + NCR_{yt,g} * LMS_{m,yt+1,g} \quad (4)$$

where  $Z$  is the adjusted labor force statistics for specific group  $g$  in year  $y$  and month  $t$ . We are mainly interested in two labor market outcomes: unemployment rate and employment-to-population ratio. Thus, two sets of calculations are performed. First,  $CR_{yt,g}$  is the coverage ratio for year  $y$ , month  $t$ , and demographic group  $g$ , and  $NCR_{yt,g}$  is the fraction of a specific demographic group that is not covered for year  $y$  and month  $t$ .  $LMS_{obs,yt,g}$  indicates the unemployment rate for group  $g$  in the observed data at year  $y$  and month  $t$ , while  $LMS_{m,yt+1,g}$  is the newly estimated unemployment rate for the missing observations in group  $g$ , at year  $y$  and month  $t+1$ . Furthermore, we replicate the same calculation for the adjusted employment rate and labor force participation rate. For instance, we substitute the observed rates and imputed unemployment rates for the missing observations with the observed employment rates and imputed employment rates for each specific year-month group, to calculate the adjusted employment-to-population ratios.

## Results

We calculate the percent of different transitions between one's employment status in the current month and in the following month. We denote the probability of each labor force state, including employment, unemployment, and not in the market, with  $E$ ,  $U$ , and  $N$ , respectively. Taking into account that someone may be missed in the next period, each individual's labor market transition between two consecutive months is reported, for example, as  $U \rightarrow E$ ,  $U \rightarrow U$ ,  $U \rightarrow N$ , and  $U \rightarrow M$ , which indicates the probability of transition from  $U$  at time  $t$  to  $E$ ,  $U$ ,  $N$ , or  $M$  (missing) at time  $t+1$ . Similarly, respondents making a transition from being employed at time  $t$  to any of the following statuses  $E$ ,  $U$ ,  $N$  and  $M$  are denoted as  $E \rightarrow E$ ,  $E \rightarrow U$ ,  $E \rightarrow N$ , and  $E \rightarrow M$ .

---

<sup>8</sup> as share of the labor force.

<sup>9</sup> We reported the average coverage ratio (2003–2019) for each specific group of interest in the paper. The monthly coverage rate for each race-gender-age grouping is available upon request.

The top panel of Table 1 reports the average fraction of individuals switching from being employed to any of the labor market states over 2003 and 2019 period. In addition to the overall estimates, the table is stratified by key subgroups of interest, such as race/ethnicity and age. Black respondents appeared less likely to have employment in two consecutive months, at 87.6 percent, compared to 91 percent sample-wide. Hispanic interviewees are slightly better off but were still 2 percentage points lower (88.8 percent) than average in terms of being consistently employed. Turning to the nonresponses (E→M), employed Blacks, Hispanics, and younger respondents are all disproportionately likely to be missed in the next month’s survey, with the difference between each group and the average being 1 to 3 percentage points. Indeed, it is somewhat striking that only the younger group (between ages 16 and 24) is far less likely to remained employed in the next month, at nearly 8 percentage points lower than the overall average, and is more likely to drop out of the survey or report not in the labor market.

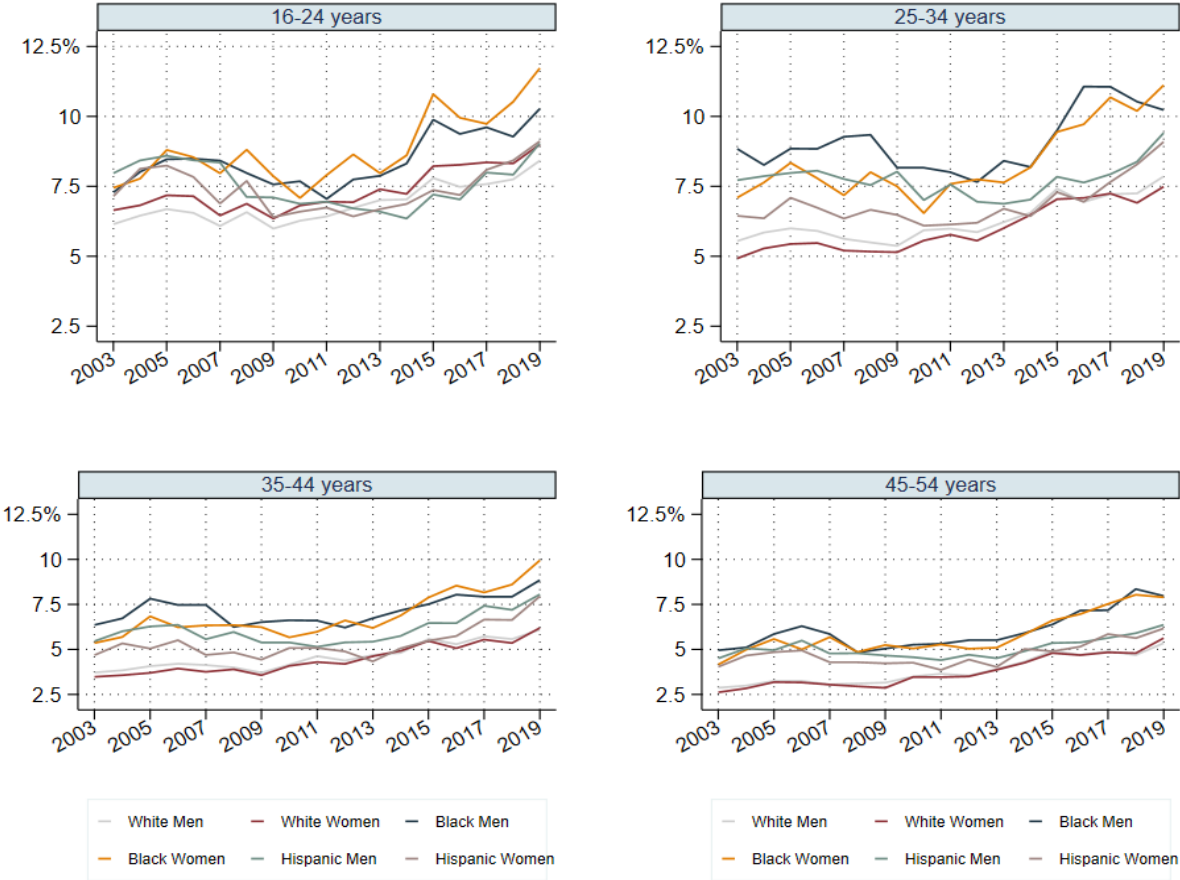
Table 1. Proportion of individuals making month-to-month labor market transition

	Total	Black	Hispanic	White	16-24	25-34	35-44
E->E	0.909	0.876	0.888	0.919	0.834	0.876	0.888
E->U	0.012	0.018	0.017	0.010	0.023	0.018	0.017
E->N	0.028	0.035	0.033	0.025	0.067	0.035	0.033
E->M	0.051	0.071	0.062	0.045	0.076	0.071	0.062
U->E	0.204	0.155	0.227	0.218	0.207	0.155	0.227
U->U	0.510	0.522	0.467	0.521	0.427	0.522	0.467
U->N	0.216	0.239	0.227	0.199	0.284	0.239	0.227
U->M	0.070	0.084	0.078	0.062	0.082	0.084	0.078
N->E	0.042	0.044	0.057	0.037	0.076	0.044	0.057
N->U	0.023	0.039	0.033	0.018	0.050	0.039	0.033
N->N	0.889	0.857	0.853	0.905	0.804	0.857	0.853
N->M	0.046	0.060	0.057	0.040	0.071	0.060	0.057

Note: Data are from the monthly CPS, 2003 to 2019. Sample includes civilian noninstitutional population over 16.

The breakdown of the portions of respondents transitioning from unemployment (U) or not in the market (N) to any labor force status is presented in the middle and bottom panels of Table 1, respectively. Focusing on the middle panel, with the exception of persistent unemployment (U→U) being more prevalent among prime-age adults rather than among 16- to 24-year-olds, similar patterns hold: Hispanics, Blacks, and younger groups tend to make the U→M or U→N transition more often than their white, older-aged counterparts do. Notably, despite a relatively low probability of continuous unemployment across two consecutive months for the 16- to 24-year-old group, they disproportionately experienced the U→N transition, at 28 percent, compared to 22 percent on average. This may signify that younger adults may be more likely to be discouraged workers and drop out of the market when unemployed.

Figure 1. Percent of nonresponse in the next month, by race/ethnicity and gender



Note: Data are from monthly CPS, 2003 to 2019. Sample includes civilian noninstitutional population aged 16 to 54.

Figure 1 reports the nonresponse rate in the next period for various demographic groups. Specifically, we calculate whether someone is missed in the following month if he or she appears in the present month’s survey. Similar gradients appear across different groups, but younger people appear to have higher nonresponse rates. Notably, an upward trend in the missing rate for most groups may imply that labor force statistics’ bias might be growing over time. For respondents aged 16-34, the Black–white difference in nonresponse rate is largest around 2016 and 2017, suggesting the nonresponse during this period may somewhat mask the true employment statistics. Specifically, among people 16-24 years old, the missing rates for Blacks prior to the Great Recession are comparable, with Hispanic young men, who have missing rates slightly higher than older people. Young Black men and women show increasing nonresponse rates in the survey after 2008, and these rates reached about 7 percent in 2015 and 2016. The picture for the young adults age 25 to 34 is much clearer. The Black–white gap in nonresponse rate appears to be about 2 to 4 percentage points before the 2008 recession period, before it converges a bit within a few years after the 2008 recession. However, the difference widens

again in 2015, with the Black nonresponse rate increasing at a faster rate. Similar patterns appear for other prime age groups, and it is less pronounced but still noticeable for respondents aged 35 to 54.

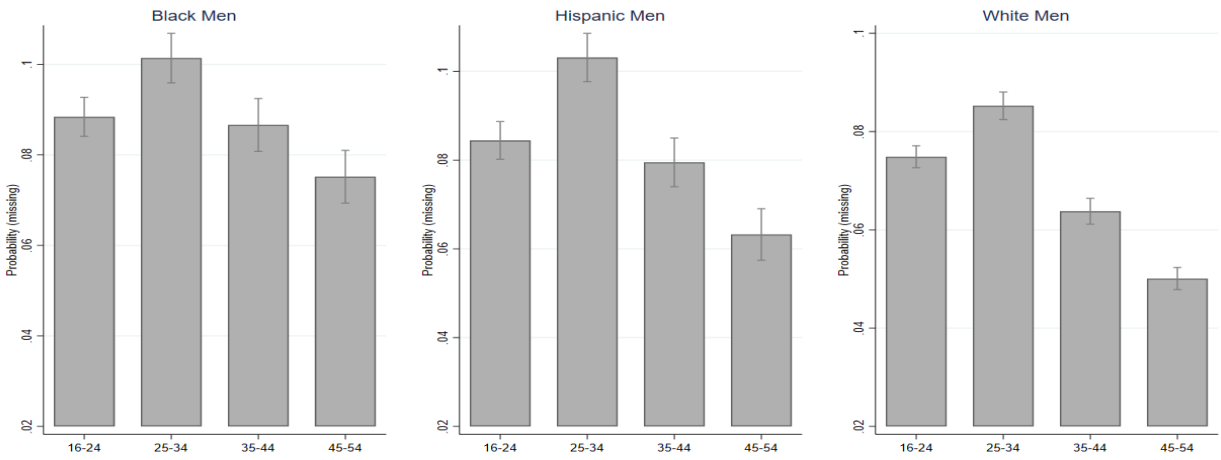
Table 2. Predicted probability of becoming nonresponse in the following month

	Model 1 Nonresponse at t+1	Model 2 Nonresponse at t+1	Model 3 Nonresponse at t+1
Employed at t	0.054	0.056	0.051
Unemployed at t	0.075	0.068	0.066
Not in the labor force at t	0.064	0.059	0.078
White, non-Hispanic		0.053	0.053
Black, non-Hispanic		0.072	0.073
Hispanic		0.062	0.059
Other		0.062	0.060
Age 16-24		0.071	0.071
Age 25-34		0.067	0.067
Age 35-44		0.050	0.051
Age 45-54		0.042	0.042
Other characteristics	No	Yes	Yes
Occupation types	No	No	Yes
Year fixed effects	No	No	Yes
State fixed effects	No	No	Yes

Note: Data are from the monthly CPS, 2003 to 2019. Sample includes civilian noninstitutional population aged 16 to 54. The estimates are based on 10,366,316 person-month observations. Other characteristics include gender and education level.

We estimate the likelihood of not responding in the next month if a person is unemployed in the current month. Table 2 presents the predicted probability obtained from a bivariate analysis and models adjusting for the respondent's age, race or ethnicity, gender, education level, job characteristics, and time and state fixed effects. As model 3 indicates, if someone is unemployed, the probability of their not responding to the survey in the following month is nearly 7 percent after a full set of covariates is introduced. Though other demographic information is included as controls, their point estimates are also of interest in understanding demographic variation in nonresponse. Consistent with the trends in Figure 1, the probability of not participating in the following survey month is highest among younger respondents (16-24), at 7.1 percent, followed by 6.7 percent for respondents aged 25-34. Approximately 6 percent of Hispanic and 7.3 percent of Black respondents do not answer the next month's survey.

Figure 2. Predicted probability of nonresponse by age and race/ethnicity



Note: Data are from the monthly CPS, 2003 to 2019. Sample includes civilian noninstitutional population aged 16 to 54. Sample for Black men includes 354,976 person-month observations; sample for Hispanic men includes 583,711 person-month observations; and sample for white men includes 2,921,594 person-month observations.

To answer the second question of the extent to which nonresponse varies by age for each racial/ethnic group, we estimate separate models for three demographic groups: Black men, Hispanic men, and white men. We regress the binary outcome of nonresponse on unemployment status, age category, and their interactions. As shown in Figure 2, unemployed Black men, ages 25 to 34, have the highest chance of not responding in the following month, over 10 percent. The next highest is Black men ages 16 to 24, who have about 8.8 percent probability of skipping the survey. In contrast, members of older age groups had a slightly lower probability of missing the next month’s survey. Hispanic men followed similar patterns, with slightly higher probability for younger groups compared to their older peers. Unemployment has a relatively lesser impact on white young adults’ nonresponse to the survey.

Above, we find that next-month nonresponse rate is clearly nonrandom in terms of one’s present labor-market activities. Although proper weighting adjustment can partially offset some of the bias resulting from missing observations, imputing values for labor-market state for the missing observations—along with considering group-specific month-to-month coverage ratios—still appears to lead to a difference in the estimate of labor-market statistics from the observed weighted values. Table 3 presents the average bias for the whole sample and decomposes the overall bias for specific subgroups. While the subsequent tables are structured by three labor market outcomes—unemployment rate, employment rate, and labor force participation rate—we focus on the discussion surrounding the unemployment and employment rates. Between 2003 and 2019, the general coverage ratio was about 87 percent, but Blacks and Hispanics were far more likely to have undercoverage issues, with coverage ratios of only 72 percent and 83 percent, respectively. When looking further into gender, Black men only had a coverage rate of 70 percent, while white women had a coverage rate over 90 percent. Black women and Hispanic

men also had low coverage rates (74% and 79%, respectively), but to a lesser extent as compared to Black men.

Table 3. Adjusted labor market outcomes by subgroups

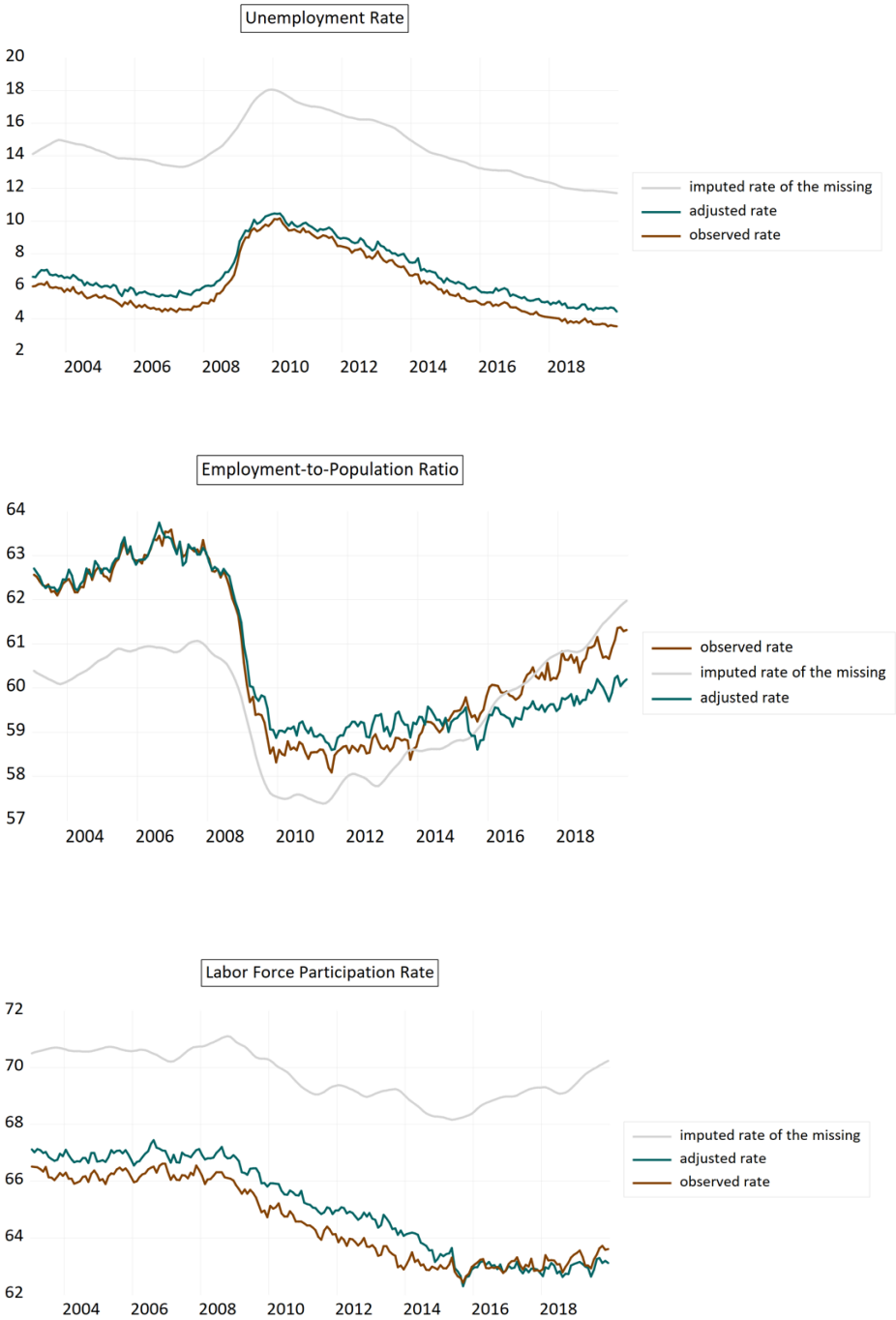
	Total	Black	Hispanic	White	Black		Hispanic		White	
					Men	Women	Men	Women	Men	Women
Coverage ratio	0.87	0.72	0.83	0.89	0.69	0.74	0.79	0.88	0.87	0.91
<b>Panel A: Unemployment rate</b> (as share of labor force)										
Observed (obs.)	6.1	10.9	7.6	4.9	11.9	9.8	7.3	8.0	5.3	4.6
Estimated of the non-respondents	14.5	20.7	16.8	12.8	21.9	19.5	14.8	18.8	12.6	12.9
Adjusted (adj.)	6.8	13.5	9.1	5.5	14.7	12.2	8.9	9.2	6.0	5.1
Obs. – Adj.	-0.7	-2.6	-1.5	-0.6	-2.8	-2.4	-1.6	-1.3	-0.7	-0.5
<b>Panel B: Employment rate</b>										
Observed	60.6	56.1	62.1	61.2	57.5	54.7	72.3	51.9	66.9	55.6
Estimated of the non-respondents	59.6	54.1	60.0	61.1	55.1	53.2	69.6	50.3	65.7	56.4
Adjusted	60.6	54.5	61.8	61.5	55.6	53.3	71.2	52.4	66.7	56.3
Obs. – Adj.	0.05	1.6	0.3	-0.3	1.9	1.4	1.1	-0.5	0.2	-0.7
<b>Panel C: Labor force participation rate</b>										
Observed	64.6	63.0	67.2	64.4	65.3	60.7	78.0	56.4	70.6	58.3
Estimated of the non-respondents	69.8	68.3	71.8	70.0	70.5	66.0	81.7	62.0	75.2	64.8
Adjusted	65.1	63.2	68.0	65.2	65.4	61.0	78.3	57.8	71.0	59.4
Obs. – Adj.	-0.5	-0.2	-0.8	-0.8	-0.2	-0.3	-0.3	-1.4	-0.4	-1.1

Note: Data are from monthly CPS, 2003 to 2019. Sample includes civilian noninstitutional population aged 16 or older. Observed rates are weighted using the person-level weight. Coverage ratio (accessed through <https://www2.census.gov/programs-surveys/cps/techdocs>) may vary from year to year (month to month).

In panel A of Table 3, we report the weighted observed unemployment rate, the estimated unemployment rate of the missing observations at time  $t+1$ , and the adjusted rate that takes into account missing observations and the undercoverage issues. Across the sample, the unemployment rate appears to have been underestimated by about 0.7 percentage point if no considerations are made for the undercoverage. This bias is more salient for Blacks, particularly Black men, with the gap between the observed and adjusted unemployment rates being about 2.8 percentage points, versus a 0.6-percentage-point difference observed for whites. This increases the already high Black-white unemployment gap, implying a gap of 8 percentage points rather than its observed 6 percentage points. The Hispanic-white gap also increases by about 0.9 percentage points after the correction. Turning to the employment-to-population rate in Panel B, it is not entirely surprising to observe that the Black men’s employment rate falls to 55.6 percent from 57.5 percent, resulting in about 2-percentage-point bias (see panel B). This is followed by

Black women, with the employment rate overestimated by 1.4 percentage points. In contrast, most groups show relatively mild differences with respect to the labor force participation rates after such a correction. The bias is pronounced among Hispanic women, with the labor force participation rate underestimated by about 1.4 percentage points.

Figure 3. Observed and adjusted labor market outcomes, 2003 to 2019



Note: Data are from the monthly CPS, 2003 to 2019. Observed rates are weighted using the person-level weight. Time series are seasonally adjusted. The line for the imputed rate of the missing data has been smoothed.

We present imputed statistics for the missing observations, weighted observed rates, and adjusted rates between 2003 and 2019 for unemployment<sup>10</sup> and employment populations in Figures 3a and 3b and for labor force participation in Figure 3c. The imputed unemployment rate for the missing observations is quite high, averaging about 14.5 percent over time, with a peak of about 19 percent around the Great Recession. On average, the national unemployment rate is understated by 0.7 percentage points during the entire examined period, with the smallest gap being less than 0.5 percentage points around 2010 and the largest gap observed in the most recent years, at nearly 1 percentage point. The picture is slightly different with regard to the national employment rate. Between 2016 and 2019, the employment rate tends to be overestimated. Prior to that, the employment-to-population rate is understated by about 0.5 percentage points. The two lines converge somewhat between 2007 and 2009 and then become relatively parallel for several years, before crossing in 2015. Overall, the actual observed and adjusted labor force participation rates tend to be parallel between 2003 and 2015, before converging in 2016 and onwards.

Figure 4. Observed and adjusted unemployment rates, by race and gender



Note: Data are from the monthly CPS, 2003 to 2019. Observed rates are weighted using the person-level weight. Time series are seasonally adjusted. The line for the imputed rate of the missing data has been smoothed.

<sup>10</sup> As a share of labor force



Figure 5. Observed and adjusted employment-to-population rates, by race and gender



Note: Data are from the monthly CPS, 2003 to 2019. Observed rates are weighted using the person-level weight. Time series are seasonally adjusted. The line for the imputed rate of the missing data has been smoothed.

Figures 4 and 5 report the decomposition results of the bias in unemployment and employment-to-population rates, paying particular attention to the Black-white difference over time. The trend in white unemployment, regardless of gender, is quite identical to the overall picture but to a much lesser degree. Inversely, a substantial gap is observed for Black men, whose largest gap was about 5 percentage points in 2019, despite a slight convergence during the Great Recession (see Figure 4). The lines for Black women tend to move in the same direction over the whole period, with an increasing gap of around 3 percentage points in more recent years. Turning to employment-to-population rate in Figure 5, the trend in every group tends to mirror the overall trend to varying degrees, with larger gaps existing among Black respondents. Across the board, it suggests that the employment rate has been overstated, especially for Black men. This is not entirely surprising, given the aforementioned understated unemployment rate among different groups. With the exception of white women, the other three groups’ adjusted rates seem to be

lower than the observed weighted values throughout the period of study. The gap between observed and adjusted employment rates tends to grow beginning in 2015. The employment rate is more overstated for Blacks than for whites. Notably, in recent years, the employment rates for Black women and men are overestimated by nearly 3 and 4 percentage points, respectively.

Table 4. Observed, adjusted labor force statistics among Black men

	16-24	25-34	35-44	45-54	55 or older
Coverage ratio	0.71	0.64	0.72	0.78	0.89
Panel A: Unemployment rate (as share of labor force)					
Observed	24.4	12.3	8.9	7.9	7.4
Estimated of the non-respondents	35.2	21.0	17.2	17.1	17.2
Adjusted	28.0	15.3	11.1	10.0	8.4
Obs. – Adj.	-3.6	-3.0	-2.2	-2.1	-1.0
Panel B: Employment rate					
Observed	38.7	73.4	77.1	71.6	36.9
Estimated of the non-respondents	37.0	67.2	71.4	66.0	37.3
Adjusted	37.0	71.4	75.5	70.2	36.0
Obs. – Adj.	1.7	2.0	1.6	1.4	0.9
Panel C: Labor-force participation rate					
Observed	51.1	83.6	84.6	77.7	39.8
Estimated of the non-respondents	57.2	85.1	86.2	79.6	44.9
Adjusted	51.6	84.3	85.0	78.0	39.3
Obs. – Adj.	-0.5	-0.7	-0.3	-0.3	0.5

Note: Data are from the monthly CPS, 2003 to 2019. Observed rates are weighted using the person-level weight.

Table 4 illustrates the adjusted labor force statistics for Black men as compared to their weighted observed rates. Notably, the coverage rate for 25- to 34-year-old Black men is only 64 percent, or 5 percentage points lower than their average. This results in an adjusted unemployment rate of 15.3 percent, about 3 percentage points higher than the observed rate (see panel A). Even more alarmingly, although the youngest group (ages 16 to 24) has about a 71 percent coverage rate, the gap in the unemployment rate appears to be nearly 4 percentage point. This might be due largely to the higher observed unemployment rate of 24 percent, compared with 12 percent for Black male workers aged 25–34. By contrast, employment rates tend to be overstated for Black men among all age groups (panel B). The largest gap in employment rate (2 percentage points) is observed for Black men aged 25–34, with other age groups (individuals over 35) experiencing an average of 0.9 to 1.6 percentage points. In panel C, the labor force participation rate has been slightly downplayed, particularly for Black young adults. The largest gap in labor force participation rate (0.7 percentage points) is observed for Blacks aged 25–34.

## Conclusion

This paper revisits a pivotal issue pertaining to the high and rising nonresponse rate in the CPS and how the nonresponse bias would affect less advantaged workers over the 2003–2019 period. The data indicates that a worker’s prior labor-market status is significantly linked to their participation in the follow-up survey, even after adjustment for the workers’ basic demographic characteristics. This finding indirectly confirms one of the explanations with respect to the rotation group bias in unemployment rate that Krueger et al. (2017) found in their work, in which they suggest that rising survey nonresponse or other factors related to noninterview likely accounts for rotation group bias.

Our results further reveal that unemployment in the current period is related to nonresponse in the following period, and the magnitude is relatively larger for Black and Hispanic workers. After imputing the nonresponse observations and adjusting for the undercoverage issue, the average unemployment rate would have been 0.7 percentage point higher than weighted observed rate, and the national employment-to-population rate appears to be 0.05 percentage point lower from 2003 to 2019, with Black men’s employment rate being overstated by about 1.9 percentage points. Despite a relatively smaller magnitude, the results are more or less consistent with recent work from Ahn and Hamilton (2020), in which the authors point out the understated official unemployment rate and suggest that the true percentage might be about 2 points higher prior to the pandemic. Furthermore, we find gaps in the unemployment rate for Black and Hispanic workers of 2.6 and 1.5 percentage points, respectively. The gaps are more alarming for younger racial/ethnic minority workers, with the gap being about 3.3 percentage points for Black men aged 16 to 34. These figures are arguably a lower boundary of the impact of nonresponse in CPS, as we do not address misclassification issues in the present paper. Also, if people who are missing for multiple observations (or never in the survey to begin with) are less likely to be employed than people who are missing for a single observation, this analysis would understate the impact of the undercoverage rate in the CPS. Next, we will further decompose results by education level, paying particular attention to less advantaged younger adults with lower education, and consider how nonresponse impacts their labor market outcomes in the long run.

In general, undercoverage is more of a problem for Blacks than for other races or ethnicities and is higher for men than for women (U.S. Census Bureau 2019). Yet, the existing survey-weighting procedure only mitigates some degree of bias introduced by the missed units and observations. The evidence presented in this paper unpacks how the missing observations are not random with respect to one’s prior labor-market state, suggesting that unemployed participants are more likely to become nonsurveyed. Notably, our estimates are based on the assumption that those who are not covered do not differ from those surveyed in ways other than race/ethnicity, age, and gender. The results may be slightly biased if coverage in the survey depends on other factors not mentioned above or on any unobserved elements. Researchers should keep this in mind when interpreting the findings.

With this in mind, the results underline the importance of the rising nonresponse rate in the CPS, which is drawing increasing attention from researchers, advocates, and policy makers. In light of the debate on full employment, despite a higher rate of labor force participation after the adjustment, we still observe a noticeably understated unemployment rate. This also raises the question of the accuracy of (un)employment statistics for Black workers, a group that has historically faced poor economic prospects. Despite the long recovery in the aftermath of the Great Recession, an increasingly alarming racial divide and inequity issues over recent years have created doubt over whether (or how much) workers from disadvantaged groups could benefit from this economic upswing. Furthermore, the findings call for investigations that are more careful with regard to survey-nonresponse issues and that explore alternative imputation approaches to offset biases from nonrandom nonresponse.

## References:

- Ahn, Hie Joo, and James D. Hamilton. *Measuring labor-force participation and the incidence and duration of unemployment*. No. w27394. National Bureau of Economic Research, 2020.
- Bailar, Barbara A. "The effects of rotation group bias on estimates from panel surveys." *Journal of the American Statistical Association* 70, no. 349 (1975): 23-30.
- Dziura, James D., Lori A. Post, Qing Zhao, Zhixuan Fu, and Peter Peduzzi. "Strategies for dealing with missing data in clinical trials: from design to analysis." *The Yale journal of biology and medicine* 86, no. 3 (2013): 343.
- Krueger, Alan B., Alexandre Mas, and Xiaotong Niu. "The evolution of rotation group bias: Will the real unemployment rate please stand up?" *Review of Economics and Statistics* 99, no. 2 (2017): 258-264.
- Meyer, Bruce D., Wallace KC Mok, and James X. Sullivan. "Household surveys in crisis." *Journal of Economic Perspectives* 29, no. 4 (2015): 199-226.
- Rothbaum, Jonathan, and Adam Bee. "Coronavirus Infects Surveys, Too: Nonresponse Bias During the Pandemic in the CPS ASEC." Social, Economic, and Housing Statistics Division Working Paper 2020-10 (2020).
- Rubin, Donald B. *Multiple imputation for nonresponse in surveys*. Vol. 81. John Wiley & Sons, 2004.
- Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles and J. Robert Warren. *Integrated Public Use Microdata Series, Current Population Survey: Version 8.0* [dataset]. Minneapolis, MN: IPUMS, 2020. <https://doi.org/10.18128/D030.V8.0>
- Schmitt, John, and Dean Baker. "Correcting Employment Rates in the 2000 Decennial Census using information from the CPS Census 2000 Match." Center for Economic and Policy Research Briefing Paper (2005).
- Schmitt, John, and Dean Baker. "The impact of undercounting in the Current Population Survey." Center for Economic and Policy Research Briefing Paper. [http://www.cepr.net/documents/cps\\_declining\\_coverage\\_2006\\_08.pdf](http://www.cepr.net/documents/cps_declining_coverage_2006_08.pdf) (2006).
- U.S. Census Bureau. *Current Population Survey, Technical Documentation, August 2019*. <https://www2.census.gov/programs-surveys/cps/techdocs/cpsaug19.pdf> (2019)
- Ward, Jason M., and Kathryn Anne Edwards. "Statistics in the Time of Coronavirus: COVID-19-related Nonresponse in the CPS Household Survey." Available at SSRN 3692799 (2020).

## Appendix

Table A1.

	A: Labor market outcomes on average		B: Labor market outcomes of the missing observations at t+1	
	without imputation	with multiple imputation	single imputation	multiple imputation
Employment-Population ratio	60.6	60.7	62.1	59.7
Unemployment rate	6.1	6.2	7.9	14.3
Labor force participation rate	64.5	64.7	67.5	69.7

Note: Data are from monthly CPS, 2003 to 2019. Sample includes civilian noninstitutional population aged 16 or older.

Table B1. Labor market outcomes with single imputation

	Total	Black	Hispanic	White	Black		Hispanic		White	
					Men	Women	Men	Women	Men	Women
Coverage ratio	0.87	0.72	0.833	0.893	0.698	0.742	0.789	0.877	0.872	0.913
<b>Panel A: Unemployment rate (as share of labor force)</b>										
Observed (obs.)	6.1	10.9	7.6	4.9	11.9	9.8	7.3	8.0	5.3	4.6
Estimated of the non-respondents	8.6	13.1	9.9	6.7	14.6	11.6	9.6	10.2	7.3	6.2
Adjusted (prj.)	6.1	11.3	8.0	4.9	12.5	10.2	7.8	8.1	5.3	4.5
Obs. – Adj.	-	-0.5	-0.4	-	-0.6	-0.4	-0.5	-0.1	-	0.1
<b>Panel B: Employment rate</b>										
Observed	60.6	56.1	62.1	61.2	57.5	54.7	72.3	51.9	66.9	55.6
Estimated of the non-respondents	62.2	57.6	62.4	63.5	57.8	57.4	71.3	53.4	67.7	59.3
Adjusted	60.9	55.5	62.2	61.7	56.5	54.5	71.6	52.8	67.0	56.6
Obs. – Adj.	-0.3	0.6	-0.1	-0.5	1.0	0.2	0.7	-0.9	-0.1	-1.0
<b>Panel C: Labor force participation rate</b>										
Observed	64.5	63.0	67.2	64.4	65.3	60.7	78.0	56.4	70.6	58.3
Estimated of the non-respondents	67.7	66.3	69.2	68.1	67.7	64.9	78.9	59.5	73.1	63.2
Adjusted	64.8	62.6	67.6	65.0	64.6	60.7	77.7	57.5	70.7	59.3
Obs. – Adj.	-0.3	0.4	-0.4	-0.6	0.7	0.02	0.3	-1.1	-0.2	-1.0

Note: Data are from monthly CPS, 2003 to 2019. Sample includes civilian noninstitutional population aged 16 or older. Under this approach, we do not consider nonresponse at time t+2 or a missed second time; therefore, if a person appeared at t and was missed at t+1, we only retain the first two labor market outcomes ( $Y_t$ ,  $Y_2$ ) in our sample post single imputation. Coverage ratio (accessed through <https://www2.census.gov/programs-surveys/cps/techdocs>) may vary from year to year (month to month).