

# Harmonized Population and Labor Force Statistics\*

John Coglianesi<sup>†</sup>, Seth Murray<sup>‡</sup>, and Christopher J. Nekarda<sup>§</sup>

Board of Governors of the Federal Reserve System

July 2025

## Abstract

The official labor force statistics often exhibit discontinuities in January, when updated population estimates are incorporated into the Current Population Survey (CPS) for the current year but are not revised backward through history. We construct harmonized population estimates spanning five decades and produce new weights for the CPS microdata that are benchmarked to these estimates. Using these weights, we estimate harmonized labor force statistics that reflect the latest available information about the population and its characteristics. The harmonized labor force series are free from the discontinuities in the historical data and show a notably larger labor force shortfall in the post-pandemic period.

JEL codes: C8, E24

Keywords: population, labor force, employment, unemployment, immigration, CPS

---

\*The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors of the Federal Reserve System. We thank Stephanie Aaronson, Jed Kolko, and John Stevens for helpful discussions and comments on earlier drafts and Chris Karlsten for superb editorial assistance. We are especially grateful to Census Bureau and Bureau of Labor Statistics staff for providing helpful comments and discussions and detailed answers to our questions.

<sup>†</sup>ORCID: 0009-0002-7660-8306. Email: [john.m.coglianesi@frb.gov](mailto:john.m.coglianesi@frb.gov).

<sup>‡</sup>ORCID: 0000-0002-7315-2895. Email: [seth.m.murray@frb.gov](mailto:seth.m.murray@frb.gov).

<sup>§</sup>ORCID: 0000-0002-2301-2720. Email: [christopher.j.nekarda@frb.gov](mailto:christopher.j.nekarda@frb.gov).

## 1. Introduction

The Bureau of Labor Statistics (BLS) publishes timely statistics about the U.S. labor market that receive wide attention each month. Many of these indicators, including the unemployment rate and the labor force participation rate (LFPR), are derived from the Current Population Survey (CPS), a monthly survey of about 60,000 households. Statistics from this survey can capture the overall labor market because individual survey responses are weighted to be representative of the demographic and geographic composition of the U.S. population. However, since the population cannot be counted in real time (except for once every 10 years in the decennial census), the weights are constructed to match population *estimates*, which can revise when underlying source data are updated.

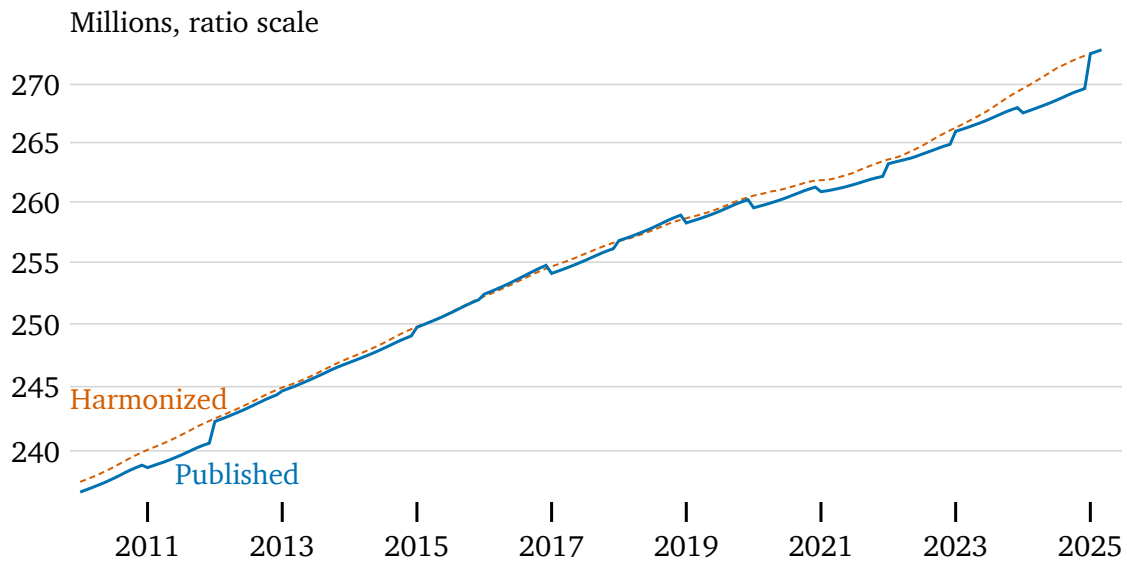
Each January, the BLS incorporates updated population estimates into the CPS for the current year, but does not revise the official household survey estimates back in history. When the revisions shift the composition of the population across demographic groups whose labor market outcomes differ, such a shift can result in large discontinuities between the updated labor market statistics for the current year and the out-of-date statistics for previous years, potentially confounding statistical analyses and assessments of the labor market. This issue affects not only the official statistics published by the BLS, but also statistics that researchers calculate from CPS microdata.

In this paper, we introduce a methodology for estimating CPS-based labor force statistics that are “harmonized”—made comparable over time—over five decades to consistently reflect the latest available data on the population. Our approach involves assembling harmonized population data at the demographic group level, reweighting CPS microdata to match these targets, and then computing time-series estimates from the reweighted microdata. Since our method closely follows the BLS’s estimation process, our harmonized labor force statistics can be interpreted as close approximations of the values that the BLS would have produced if it had been able to use the latest population data when it originally published its statistics.<sup>1</sup> We

---

1. Our approach uses almost all of the same steps as the BLS’s procedure, but it is important to note two exceptions that limit our ability to perfectly reproduce the estimates the BLS would have produced with the same population targets. First, we do not adjust

**Figure 1. Civilian noninstitutionalized population aged 16 years or older**



Source: Bureau of Labor Statistics (BLS); authors' calculations using data from the Census Bureau and the BLS.

provide harmonized estimates for the unemployment rate, the LFPR, and other labor force statistics, along with harmonized microdata weights that researchers can use to reproduce any CPS statistic adjusted for population revisions. We plan to update both the microdata weights and the harmonized time series annually to reflect each new vintage of population estimates from the Census Bureau.

By not revising the historical time series of labor force statistics to reflect new population data, the BLS's standard practices often introduce discontinuities in the time series between each December to January. These discontinuities are evident in the published series for the civilian noninstitutionalized population (CNP) aged 16 years or older, shown by the solid blue line in figure 1. For example, in January 2012, the CPS incorporated population estimates derived from the 2010 Census, which resulted in a jump of 1.7 million from December 2011 to January 2012, even though the typical pace of change before and after that period was only about 200,000 per

---

weights in the state coverage step and state-level targets in second-stage weighting to reflect the latest population estimates since the population targets for these steps are not publicly available. Second, we base our estimates on the public-use microdata files, which in some years feature slightly perturbed demographic information for confidentiality reasons, while the BLS uses unperturbed data for its estimates.

month. Similarly, the revision in January 2025 was particularly large, revising up the population level by nearly 3 million. Population revisions can also be negative, such as in 2017, 2019, and 2020.

Our harmonized time series, shown by the dashed orange line in figure 1, does not exhibit any such discontinuities because all years of the harmonized series reflect the latest population data. Our approach to constructing the harmonized series involves several steps.

We begin by assembling Census Bureau estimates for the population at the demographic subgroup level for the period in between each decennial census and adjust for the “surprise” contained in the actual decennial census count. Harmonizing these detailed population estimates involves some adjustment between eras in which data were collected differently, but maintaining the fine level of demographic detail is essential for our approach. We harmonize population estimates from January 1976 through April 2020, and for the subsequent period we use the Census Bureau’s latest published estimates—currently the Vintage 2024 (V2024) estimates, which were released June 26, 2025.

With these estimates in hand, we replicate the BLS’s weighting and estimation procedures to produce time series with properties similar to those of the official series. We construct new CPS microdata weights at broadly the same level of demographic detail that the BLS uses while ensuring that these weights match our harmonized population estimates—and thus reflect the latest population data. We also replicate the method that the BLS uses for producing composite time-series estimates that exploit the longitudinal structure of the CPS. Using published seasonal factors and adjusting for the CPS survey redesign, our time-series estimates replicate as closely as possible what the BLS would have produced if it had our current harmonized population estimates in real time. As a test of the fidelity of our replication, we are able to reproduce the statistics published by the BLS that show how revisions to the current-year population estimates would have affected the preceding year’s December labor market statistics. Although our estimates are nearly identical, we cannot perfectly reproduce the BLS’s estimates, because the publicly available CPS microdata have perturbed age information in order to protect the confidentiality of respondents.<sup>2</sup>

---

2. See section 4.2 for further discussion.

As described in section 5, the time series we produce are smoother than official estimates and feature some notable differences, particularly in the period around the pandemic. For instance, our estimate of the LFPR is above the published estimate in February 2020, on the eve of the pandemic, and shows a slower, but more even pace of recovery during 2021–22. Our estimates indicate that the labor force shortfall in the post-pandemic period was 1.5 million larger than published data indicated.

Although our main estimates begin in 1976, when CPS microdata first become available, we also extend key time series back to 1948. Our harmonized series smooth out several substantial breaks in the time series that resulted from the introduction of updated population controls drawn from the latest decennial census. All told, our harmonized time series cover 1948–2024 for the major labor force statistics and demographic groups.

In addition to headline series, we can construct estimates of any CPS statistic on a harmonized population basis. Since we produce new CPS microdata weights consistent with harmonized population targets, researchers can correct the statistical series that they derive from the microdata for the effects of population revisions simply by using our weights in place of the published weights. We outline an example case—comparing trends in prime-age native-born population for men and women—to demonstrate the value of estimates corrected for population revisions.

Over the past several years, immigration has contributed more to population growth than published data previously indicated. As a result, when the latest official estimates were released, the population aged 16 years or older at the end of 2024 was revised up by nearly 3 million people (figure 4). Before this revision was implemented, the large contribution of immigration was evident from administrative records on migrant flows, and many observers noted at the time that population growth was likely much more rapid than official estimates indicated (see, e.g., Congressional Budget Office 2024). We show how our method could have been applied before the publication of the revised official estimates. We construct an alternative set of population estimates that incorporates an adjustment at the detailed demographic group level using data on migrant flows, and then we apply our methodology to these estimates. These immigration-adjusted labor force estimates are very close to the latest official statistics, correctly indicating

that the LFPR revised up in recent years while the unemployment rate was essentially unaffected. This exercise highlights that our methodology does not inherently rely on official population estimates, and can be used with *any* set of population estimates with sufficient demographic detail.

The new estimates we produce help shed light on the post-pandemic labor market recovery. Previous studies analyzing the pace of labor market recovery and the extent of recovery remaining have tended to rely on either the BLS's published series or author-constructed series based on published CPS microdata, both of which have been subject to large breaks in post-pandemic years (Cooper et al. 2021; Forsythe et al. 2022; Hobijn and Şahin 2022). Robertson (2023) and Robertson and Willis (2022) represent notable exceptions, as they adjusted published series for population revisions by assuming the revisions phased in smoothly since the base period. Our approach instead allows the underlying source data to dictate how the population revisions are distributed since the previous base period, thus accounting for potentially nonmonotonic revisions over history.

Our new microdata weights make it easy for researchers to produce statistics consistent with our harmonized population, because our weights can be used as a drop-in replacement for the published weights. These harmonized microdata weights contribute to the previous literature that has enhanced the utility of the CPS microdata, such as with longitudinal linking procedures (Drew, Flood, and Warren 2014; Madrian and Lefgren 2000), wage cleaning (Schmitt 2003), and harmonized industry and occupation codes (Flood et al. 2024), among other innovations.

Our methodology can readily produce labor market statistics that are consistent with alternative population estimates, thus enabling analysis and discussion about how large and fast-changing population dynamics affect the labor market. Edelberg and Watson (2024) use alternative population estimates from the Congressional Budget Office (CBO) to argue that this immigration surge resulted in much faster growth in employment and the labor force over 2022–24 than the BLS's published statistics showed. Similarly, the BLS (2025) recently published a set of experimental time-series measures that smooth the December 2024 revisions to the population, labor force, and employment back to April 2020. More closely related to this paper, Kolko (2025) produces microdata weights consistent with alternative

population estimates that smooth the revisions to population at the end of 2024 back to April 2020.

The rest of the paper proceeds as follows. Section 2 describes inconsistencies in the microdata weights over time. Section 3 provides a high-level overview of the BLS’s process for constructing time-series estimates. Section 4 details how we construct harmonized population estimates and replicate the BLS’s process to produce our own time-series estimates. Section 5 reports our new estimates of headline series, showing how they differ from published estimates.

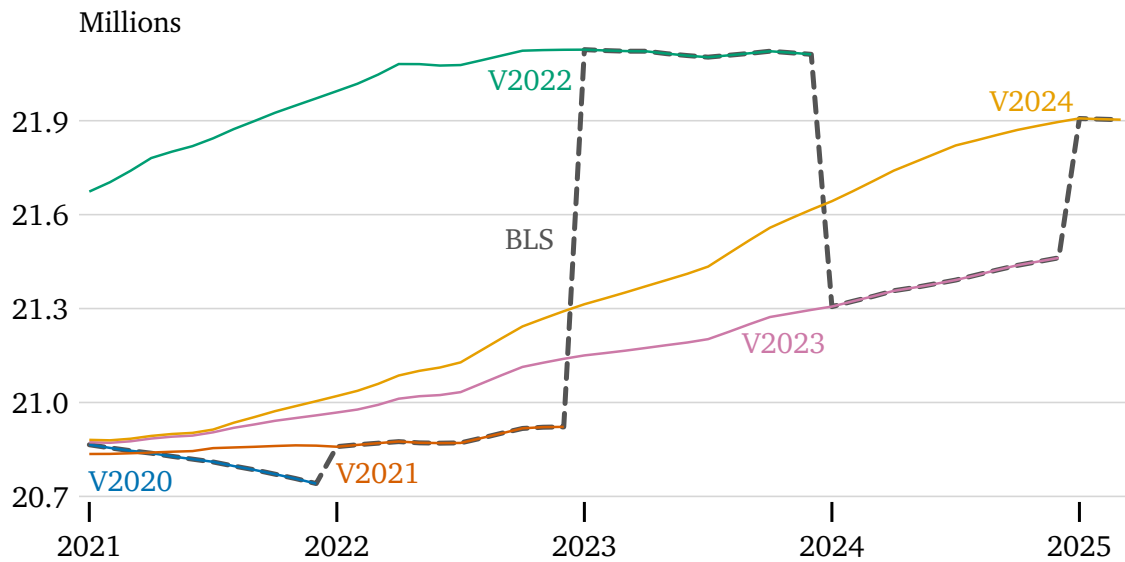
## 2. Population revisions in the CPS

CPS microdata are released each month, with the weights constructed to target the estimated population level for that month from the Census Bureau’s then-current vintage of population estimates—ensuring that labor force statistics constructed from the microdata files are consistent with the then-current population estimate. Before the start of each calendar year, the Census Bureau produces a new vintage of population estimates using data available as of the previous June. While these population estimates represent the Census Bureau’s best projection as of the start of the year, the population estimate targeted by the CPS sample weights for a month in that year is based on source data that may be up to 18 months out of date.

Because the CPS files are typically not revised, the weights for a month will always add up to the population estimate that was current at the time, even if the Census Bureau later revises its population estimate for that month. The black line in figure 2 shows the BLS’s published population for individuals aged 20 to 24, with the colored lines showing different vintages of the Census Bureau’s population estimates for this group. Within each calendar year, the BLS series matches the then-current Census Bureau population estimates. At the start of each year, the CPS estimates are controlled to a new vintage of population estimates, resulting in discontinuities between December and January—even though the population estimates from the Census Bureau are smooth across years (within any given vintage).

The key implication of this pattern is that CPS weights for all years except the current year are outdated, since they reflect older population estimates

**Figure 2. Civilian noninstitutionalized population aged 20 to 24**



Source: Bureau of Labor Statistics; Census Bureau.

that have since been superseded by newer estimates. The outdated weights will not generally add up to the latest population estimates, either in aggregate or for demographic subgroups, and the distance between the sum of the weights and the latest population estimates can be quite large.

Moreover, the difference between published CPS weights and the latest population estimates may have an alternating sign over time, with positive deviations in some months and negative deviations in others. For example, the orange line in figure 2 shows the Census Bureau's latest estimate of the population aged 20 to 24, which is designated as the V2024 series. This V2024 estimate of the population aged 20 to 24 is higher throughout history than the Vintage 2020, Vintage 2021, and Vintage 2023 (V2023) series, but lower than the Vintage 2022 series. As a result, the published BLS series (the black dashed line) is below the V2024 estimate in 2021–22 and 2024, but above it in 2023. Although this example focuses on a single age group over a short period, the general pattern of alternating-signed revisions is not uncommon when comparing the BLS's published population series with the latest Census Bureau estimates.

These patterns make it clear that merely smoothing the BLS series is not sufficient to recover the correct up-to-date population estimates. One com-

monly suggested idea for addressing the large discontinuities each January is to simply take the jump for each demographic group and smooth it back linearly in time to some base period (e.g., the previous January, or the previous decennial census date). However, this approach will not correctly account for groups where the slope of the population path changes in newer vintages, as shown earlier, and, more generally, is not going to reproduce population estimates that match the latest Census Bureau estimates. It is also incorrect to apply a time-series smoothing filter to the population to address the January jumps, since doing so merely spreads out the errors across the year but does not undo them.

Although these revisions only apply directly to population, they will indirectly affect labor force statistics derived from the CPS to the extent that the statistic of interest varies across groups that revise significantly. For example, the latest Census Bureau population estimates for April 2020 feature nearly 1 million more men aged 25 to 54 and nearly 1 million fewer women aged 55 years or older, than does the BLS series. The former group tends to have higher-than-average participation rates, and the latter group tends to have lower-than-average participation rates, so both of these differences imply that adjusting the population to match the latest Census Bureau estimates will raise the LFPR relative to the BLS's published estimate. In this way, the issue is not just a matter of outdated population estimates in the microdata, but may affect *any* statistic computed from the CPS.

### 3. Overview of the BLS' methodology

In this section, we provide a high-level overview of the BLS's procedure for constructing time-series estimates of labor force statistics. This procedure involves four main stages to go from individual-level responses in the CPS to nationally representative aggregate time series. Our summary draws heavily on *Current Population Survey Design and Methodology* Technical Paper 77 (U.S. Census Bureau 2019, hereafter "CPS Technical Paper 77"), to which interested readers should refer for more information.<sup>3</sup>

1. **Weighting based on geography:** The first stage constructs a sampling weight for each individual based almost solely on geographic

---

3. Especially interested readers may also wish to review earlier editions of the CPS Technical Paper series, such as U.S. Census Bureau (2002, 2006)

information. This weight is the product of three factors: (1) a “base weight” reflecting the probability of sample inclusion, (2) a “nonresponse adjustment factor”, which redistributes weight from nonresponding units to responding units within a narrow geographic cell, and (3) a “first-stage adjustment factor”, which accounts for differences in the Black population between geographic units included in that month’s sample relative to state-level totals. Importantly, each of these factors either is the same for all individuals within a geographic cell (as the first and second terms are) or can be calculated in advance and are independent of survey responses (as the first and third terms are).

2. **Reweighting based on demographics to match population estimates:** This stage reweights demographic groups within the sample so that subgroup population totals match Census Bureau estimates. Sampling weights are adjusted to match detailed national race/ethnicity/sex/age targets (“national coverage step”), then to match state-level race/sex/age targets (“state coverage step”), and then, finally, to jointly match three coarser sets of population targets (“second-stage weighting”). This final adjustment involves using iterative proportional fitting (or “raking”) to jointly match national race/sex/age targets, national ethnicity/sex/age targets, and state/sex/age targets. See CPS Technical Paper 77 for the exact definitions of demographic groups used in each of these population estimates.

The demographic group-specific population estimates used in this phase come from the Census Bureau’s Population Estimates Program (PEP). Each year, this program releases updated monthly estimates of the civilian noninstitutionalized population for detailed demographic groups, based on source data measuring births, deaths, immigration, and other flows. These estimates cover the period since the most recent decennial census and include a projection through the current year. The BLS aggregates the population across these detailed demographic groups into the more aggregated demographic groupings used as targets for each step.

3. **Calculating composite estimates:** The BLS’s estimates for not seasonally adjusted (NSA) time series such as employment and unem-

ployment each month are a composite of two separate estimates. The first estimate is the quantity total, summed up using the second-stage weights constructed in the previous phase. The second estimate uses the longitudinal nature of the CPS, taking the month-over-month change in the quantity among continuing respondents and adding to the previous month's composite estimate. Huang and L. R. Ernst (1981) show that constructing a composite estimate that incorporates both stock and flow information in this way reduces the variance of time-series estimates.

For ease of reproducing these estimates, the BLS adds to the CPS a set of "composite weights", which add up to the composite estimates for a given month at the demographic group level. These composite weights are produced by raking weights within state, national race/sex/age, and national ethnicity/sex/age cells such that total employment and unemployment within the cell matches the composite estimate.

**4. Seasonal adjustment, including adjustment for known breaks:**

The final stage involves adjusting the NSA series for seasonality and known discontinuities. The BLS uses X-13ARIMA-SEATS to conduct seasonal adjustment, outlier detection, and adjustment for level shifts that occur on known break dates. These known breaks include the dates when updated population estimates were introduced, as well as the introduction of the CPS redesign in 1994 and several other CPS changes (see CPS Technical Paper 77, chapter 2-5, for a complete list).

To produce headline series, such as the unemployment rate or the LFPR, the BLS seasonally adjusts component series and then calculates the headline series from the seasonally adjusted components. For example, the seasonally adjusted labor force level is the sum of eight separately seasonally adjusted components: the levels of unemployment and employment each among four demographic groups (men and women/aged 16 to 19 and aged 20 years or older). The seasonally adjusted unemployment rate is the ratio of total unemployment (sum of seasonally adjusted unemployment in each of the four demographic groups) to the seasonally adjusted labor force level.

Each January, the BLS incorporates a new vintage of population estimates for the current year from the Census Bureau, which it uses as the population estimates in the second stage, as described previously. Although these population estimates extend back to the previous decennial census, previous years' weights are not revised, nor are the aggregate time-series estimates based on the CPS.

One might wonder how the aggregate time-series estimates would differ if the BLS revised previous years' data to reflect the most up-to-date population estimates. The key phase where population estimates enter the process for calculating time-series estimates is the second phase, in which observations are reweighted based on demographics. If the BLS had different population estimates during this phase, it would lead to different second-stage weights in the microdata, different composite estimates (and weights), and, ultimately, different estimates for aggregate seasonally adjusted time series. Importantly, different population estimates for demographic groups would not affect the first-stage weights, only the second-stage weights and resulting estimates. This aspect implies that one could replicate what the BLS would have produced if it had different population estimates by using the existing weights (which incorporate the first-stage adjustment) and replicating the reweighting and estimation steps in phases two through four. This is the essence of our approach, which we describe in the next section.

#### 4. Our methodology for harmonized estimates

This section provides a high-level overview of how we construct harmonized time-series estimates adjusted for population revisions. Full details are provided in section A of the appendix.

Our approach begins by constructing a new set of harmonized population estimates and then follows the BLS's current methodology to construct new CPS microdata weights that reflect these harmonized population estimates. With these new weights, we replicate the BLS's composite estimation and adjustment procedures to arrive at time series consistent with the latest population data. These time series closely approximate what the BLS would have published if it had the harmonized population series, which reflects the latest population data, at the time it had constructed its estimates.

#### 4.1. Harmonized population estimates

Our methodology for constructing harmonized monthly population estimates involves combining estimates across five decades, from January 1976 through April 2020 (the latest decennial census), and adjusting for differences across time. For the postcensal period, we can directly use the latest published estimates from the Census Bureau (currently V2024).

As mentioned previously, the BLS uses population estimates from the Census Bureau's PEP as external population controls in constructing the CPS weights. These monthly estimates cover the period between decennial censuses, where the estimate for a specific month starts with the population from the previous decennial census, adds births, subtracts deaths, and adds net migration.<sup>4</sup> Estimates of the population flows are based on a number of different data sources—for example, births and deaths are estimated from data collected by the National Center for Health Statistics (NCHS). However, many data sources are released only with a lag (e.g., NCHS data are about two years lagged), and so the population estimates for the most recent several years of a vintage are necessarily based on projections of some source data.

Each year, PEP releases a new vintage of estimates incorporating updated source data and projections. As a result of the incoming data and revised projections, the new vintage may differ from the previous vintage over all months back to the previous decennial census, although the largest revisions are typically in the most recent years. The new vintage will also extend one year further than the previous vintage. Once every 10 years, the new vintage will also incorporate a new decennial census and therefore will cover a much shorter period than the previous vintage. For example, the vintage released in 2011 was based on the 2000 Census and covered April 2000 through December 2011, while the vintage released in 2012 incorporated the 2010 Census and covered April 2010 through December 2012.

To form harmonized population series that reflect all available information, we combine two sets of data:

---

4. U.S. Census Bureau (2024b) contains a detailed description of the methodology for the V2024 estimates.

1. The decennial census counts are the most accurate information about the population and its characteristics—but they cover only one month every 10 years.
2. The postcensal estimates cover all months after the decennial census—but they are subject to error.

We combine these two sources, maximizing their individual strengths, resulting in harmonized population series that are free from discontinuities.

Specifically, we adjust the time series from the last postcensal estimate for a decade, which contains the best estimate of the monthly path of the population over the decade, to smooth the transition between decennial censuses, which are the most accurate counts of the population at those two points in time.<sup>5</sup> Errors between the postcensal estimate and what the census count ultimately showed can arise because the PEP’s method for estimating population stocks by accumulating flows is imperfect and leads to some projection errors.<sup>6</sup> Importantly, the ultimate postcensal error could arise from a combination of errors occurring at any points between decennial censuses, so our harmonized estimate redistributes this error evenly across the decade. These harmonized population series are constructed separately for each group defined by age, sex, race, and Hispanic origin.

An example helps clarify these ideas. Figure 3 shows our harmonized estimate of the population for 16-year-old non-Hispanic Black men along with the relevant vintages of Census Bureau estimates. The pink line is the last postcensal estimate based on the 2010 Census (the pink circle). The gold circle shows the 2020 Census value for April 2020, and the gap between the last point of the pink line and the gold circle is the postcensal error.<sup>7</sup> Our harmonized population estimate for this group, the black line, smooths the

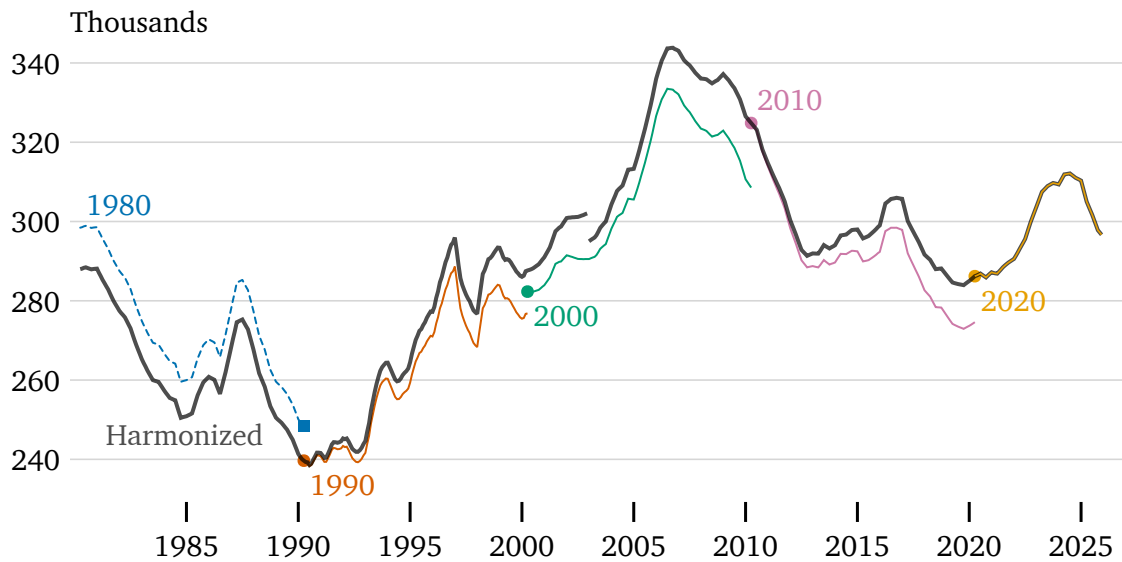
---

5. Our harmonized estimates are conceptually similar to what the Census Bureau calls “intercensal estimates”, and we use the same methodology (U.S. Census Bureau 2024a). We call our estimates “harmonized” to avoid confusion with the Census Bureau’s official data product.

6. These projection errors are discussed further in Serrato and Wingender (2016), who use them as a source of exogenous variation in federal spending to estimate local multipliers.

7. Note that the points for decennial censuses through 2010 are based on full counts, while the point for the 2020 Census comes from a “blended base”, which uses the full count for aggregate statistics but uses some estimates for determining the demographic composition. Official CNP estimates from the PEP use the blended base for April 2020, so we follow this approach in our harmonized population estimates.

**Figure 3. Population of 16-year-old non-Hispanic Black men**



Note: Solid colored lines are postcensal estimates for the civilian noninstitutionalized population; solid circles denote the decennial census value. The dashed line is the civilian population (CP); the solid square is the 1990 Census value for the CP. The break in the harmonized series in January 2003 reflects a change in the concept of race.

Source: Census Bureau; authors' calculations using data from the Census Bureau.

postcensal error back to April 2010. The gold line is the V2024 estimate, the latest postcensal estimate based on 2020 Census. Since this estimate is the best available, our harmonized series is equal to the postcensal estimate from April 2020 forward. In this way, we create a harmonized series for population that smooths through each decennial discontinuity, while using the timeliest estimate of the path of population within each decade.

There are two implementation challenges with this approach worth noting, along with how we address each one:

**Change in classification of race:** The 2000 Census was the first to allow respondents to check as many boxes as necessary to identify their race, whereas in previous censuses, responses to the race question were limited to a single category. The CPS would eventually adopt the same survey approach, but not until January 2003. As a result, the population estimates for 2000–02 are not comparable with the race categories identifiable in the CPS microdata. For example, the break in harmonized series in January 2003 shown in figure 3 reflects this change in the concept of race. We ac-

count for this change in three steps (which are described in more detail in section A4 of the appendix):

1. We take the PEP estimates released in 2011 (the last to be based on the 2000 Census) and adjust for the April 2010 decennial census wedge using the formula given earlier. This adjustment delivers estimates for 2000–02 on a multi-race basis.
2. We then convert these estimates from a multi-race basis to a single-race basis. We assume that respondents who identify as a particular race alone in the multi-race basis would identify as that race in the single-race basis. For respondents identifying as a combination of races, we assume that they would be equally likely to identify as any race in the combination under the single-race basis. This approach delivers single-race-basis estimates for 2000–02.
3. We can additionally take the single-race basis estimate for April 2000 and use it to compute the decennial census wedge for that date, comparing against the last PEP estimate for April 2000 based on the 1990 Census (which used the single-race survey question). Smoothing this wedge back gives single-race-basis estimates for 1990–2000.

The result of these steps is population series that reflect the latest estimates and are consistent with the CPS questionnaire over 1990–2002. Although we smooth through decennial census breaks, our harmonized estimates by race have a break in January 2003, when the classification changes from a single race to multiple races.

**1980–90 estimates basis:** The latest vintage of PEP estimates for 1980–90 is available only at a quarterly frequency and covers only the civilian population. To get monthly estimates for the civilian noninstitutionalized population, we make two adjustments. First, we interpolate the quarterly series to monthly. Second, we multiply each estimate by the ratio of civilian noninstitutionalized population to civilian population on April 1990, separately by demographic group.

## 4.2. Reweighting and composite estimation

With a set of consistent, harmonized population estimates in hand, we then replicate the BLS's weighting procedure. Since the published microdata weights already include the adjustments based on geography (described earlier in the first phase of the BLS weighting), we do not need to replicate this step. We take the published second-stage weights and then replicate the second phase of the BLS weighting using our harmonized population estimates.<sup>8</sup> Section B in the appendix provides additional details.

The first step in reweighting is replicating the “national coverage step”. We divide CPS respondents into demographic cells defined by the intersection of age, sex, race, ethnicity, and month-in-sample (MIS). (See table B1 in the appendix for the cell definitions.) For a demographic cell  $d$ , we modify the published second-stage weights  $w_{i,t}^{\text{Second stage}}$  of individuals in the group to match the population target  $p_{t,d}^{\text{Harmonized}}$ , giving the national-coverage-adjusted weight  $w_{i,t}^{\text{Natl. coverage}}$ :

$$(1) \quad w_{i,t}^{\text{Natl. coverage}} = w_{i,t}^{\text{Second stage}} \left( \frac{p_{t,d}^{\text{Harmonized}}}{\sum_{i \in d} w_{i,t}^{\text{Second stage}}} \right) \quad \text{for } i \in d.$$

The next step is to replicate second-stage weighting. CPS respondents are divided into two distinct sets of groups: one defined by the intersection of race, sex, age, and MIS and the other defined by the intersection of ethnicity, sex, age, and MIS. (See tables B2 and B3 in the appendix for the cell definitions). For 10 iterations, weights are updated to match each of these targets in turn. At each iteration, the previous weights  $w'_{i,t}$  are updated into new weights  $w''_{i,t}$ :

$$(2) \quad w''_{i,t} = w'_{i,t} \left( \frac{p_{t,d}^{\text{Harmonized}}}{\sum_{i \in d} w'_{i,t}} \right) \quad \text{for } i \in d$$

---

8. We note that our procedure omits the state-level weighting steps in the BLS's procedure. Between the national coverage step and second-stage weighting, the BLS implements a “state coverage step”, which matches state-level demographic totals. Also, during second-stage weighting, the BLS targets a third set of population estimates at the state/sex/age level, in addition to the race and ethnicity targets that we use in our replication. The PEP estimates that form the external population controls do not include state-level monthly estimates of the CNP by demographic group, so we are unable to directly replicate the state weighting steps with these estimates. Incorporating state-level population estimates is outside the scope of the current analysis, but this is a fruitful direction for future research.

using each of the definitions of demographic group cells  $d$ . This iterative proportional fitting, or “raking”, procedure ensures that both sets of targets are matched as well as possible (Stephan 1942).

After replicating these steps, we can produce second-stage weights consistent with our harmonized population estimates. These weights are updated at the individual level in the microdata, making it possible to reproduce any analysis in the microdata in a way that is consistent with harmonized population. Researchers can benefit from these weights in that they do not require any changes to researchers’ methodologies, only using a new weight variable instead of the published weights. For calculating monthly averages (or any other moment) in the microdata, the second-stage weights are the preferred choice of weight.

To produce time series for major labor force statistics, such as the unemployment rate or the LFPR, we replicate the BLS’s composite estimation procedure described on pages 76–78 of CPS Technical Paper 77. (See section C in the appendix for details, including cell definitions.) The BLS’s method follows Breau and L. Ernst (1983) in calculating the composite estimate for employment as a weighted average of two separate estimates: a “levels” estimate, defined as total employment among all rotation groups in the current month, and a “changes” estimate, defined as the previous month’s composite estimate plus the change in employment among rotation groups that can be matched longitudinally backward in time (i.e., MIS 2–4 and MIS 6–8).<sup>9</sup>

We verified that our procedure exactly replicates the BLS’s composite estimates when using the published second-stage weights, lending credence that our estimates using our new harmonized population estimates are consistent with the BLS’s methodology.<sup>10</sup>

---

9. The estimate also includes an adjustment for the difference in employment between incumbent and entering rotation groups. The composite estimate for unemployment is defined similarly as for employment. For both variables, we take the coefficients on each term in the composite estimate from CPS Technical Paper 77.

10. We are able to replicate the BLS’s composite estimates for most months over 2003–10, which is the period using the modern composite estimator and before the BLS began perturbing respondent information in the public-use microdata files (PUMF). [Beginning in January 2011](#), the Census Bureau incorporated additional safeguards in the CPS PUMF to ensure that respondent identifying information is not disclosed. In general, respondents’ ages were altered, or “perturbed”, in the PUMF to further protect the confidentiality of survey respondents and the data they supply. Thus, although we cannot exactly match the

An additional benefit of our replication is that we can construct composite estimates consistently for the entire period that CPS microdata are available. The BLS's current composite estimation procedure was introduced only in January 1998 and coincided with the introduction of convenience composite weights into the microdata.<sup>11</sup> We apply the BLS's current composite procedure throughout and provide microdata weights consistent with these estimates.

#### 4.3. Seasonal and other time-series adjustments

To maximize compatibility with published estimates, we use published seasonal factors instead of re-estimating seasonal factors. Our focus in this analysis is on the effect of adjusting for population revisions, and so re-estimating seasonal factors would risk complicating the differences between our estimates and published series by introducing an additional margin along which they could differ. Nonetheless, adjusting for population revisions could affect the seasonality of the series, and so re-estimating seasonal factors might deliver notably different estimates.<sup>12</sup> We view this step as a useful extension for future research to consider.

The CPS implemented a major redesign in January 1994, which changed the measurement of many of the estimates derived from the CPS. We apply adjustment factors from Polivka and Miller (1998) to our harmonized time series for 1948–93, so that the resulting series are comparable with the post-redesign measures.

Additionally, we extend key time series back to 1948, by adjusting the published series for known breaks. CPS microdata are only available starting in 1976, so for earlier years, our harmonized series are derived from the

---

BLS's published estimates after 2010, which are based on unperturbed data, our estimates are very close.

11. Under the old procedure, composite estimation was performed at the macro level, combining the second-stage estimate for the current month with the composite from the preceding month and an estimate of change from preceding to current month. Over time, the CPS refined the composite, updating the weights used in the weighted average as well as adding a component that captures the net difference between the incoming and continuing parts of the current month's sample. See U.S. Census Bureau (2002, 2019).

12. Additionally, if we re-estimated seasonal factors, we would be able to estimate factors for January in the same way as for other months, in contrast to the official estimates, which are affected by controlling for a level shift each year due to the introduction of new population estimates.

published time series. We collect information reported in historical issues of *Employment and Earnings* about the effect of the updated population controls and use it to smooth out the discontinuities in the time series for the CNP and key labor force statistics. The data and methods for harmonizing the pre-1976 time series are described in section D2 of the appendix. All told, our harmonized time series cover 1948–2024 for the major labor force statistics and demographic groups.

#### 4.4. Validating our methodology

With each January employment report, the BLS releases special tabulations of December data that incorporate the new population controls to show their effect on the labor force data. Since our methodology uses the latest population estimates, our harmonized series should match these special tabulations for December. Indeed, as shown in table D2 in the appendix, our estimates are very close to the BLS’s special tabulations for December 2024.<sup>13</sup> As a result, we are confident that our methodology reproduces the BLS’s weighting, composite estimation, and time-series adjustment steps as accurately as possible with publicly available data.

### 5. Results

Our time-series estimates based on harmonized population estimates provide new estimates of headline labor force statistics. In this section, we present several of these series and describe the important differences relative to published series. Particularly in the post-pandemic period, our series show notable differences from the published LFPR and population estimates, both overall and for specific groups. However, the unemployment rate has been less affected by revisions to population estimates.

---

13. The small differences may be due to the effects of privacy protections in the CPS microdata, particularly perturbing ages in the published microdata. The published revision table is calculated based on unperturbed microdata used internally by BLS, while our estimates use the perturbed public data.

## 5.1. Headline series

We start by presenting updated estimates for population, the LFPR, and the unemployment rate. These headline measures constructed from the CPS are among the most examined labor market indicators.

Our population estimates are notably smoother than published data. The top panel of figure 4 shows the CNP for men and women aged 16 years or older. The published versions of these series have sharp time-series breaks stemming from the introduction of new population estimates, which are not present in our series.<sup>14</sup> As shown in the lower panel, our harmonized estimates can be either above or below the published series, and these differences can persist for substantial periods (e.g., our estimate for men is more than 1 million above the published series for most of 2017–21). Moreover, the differences vary by characteristics, and not necessarily in offsetting ways.

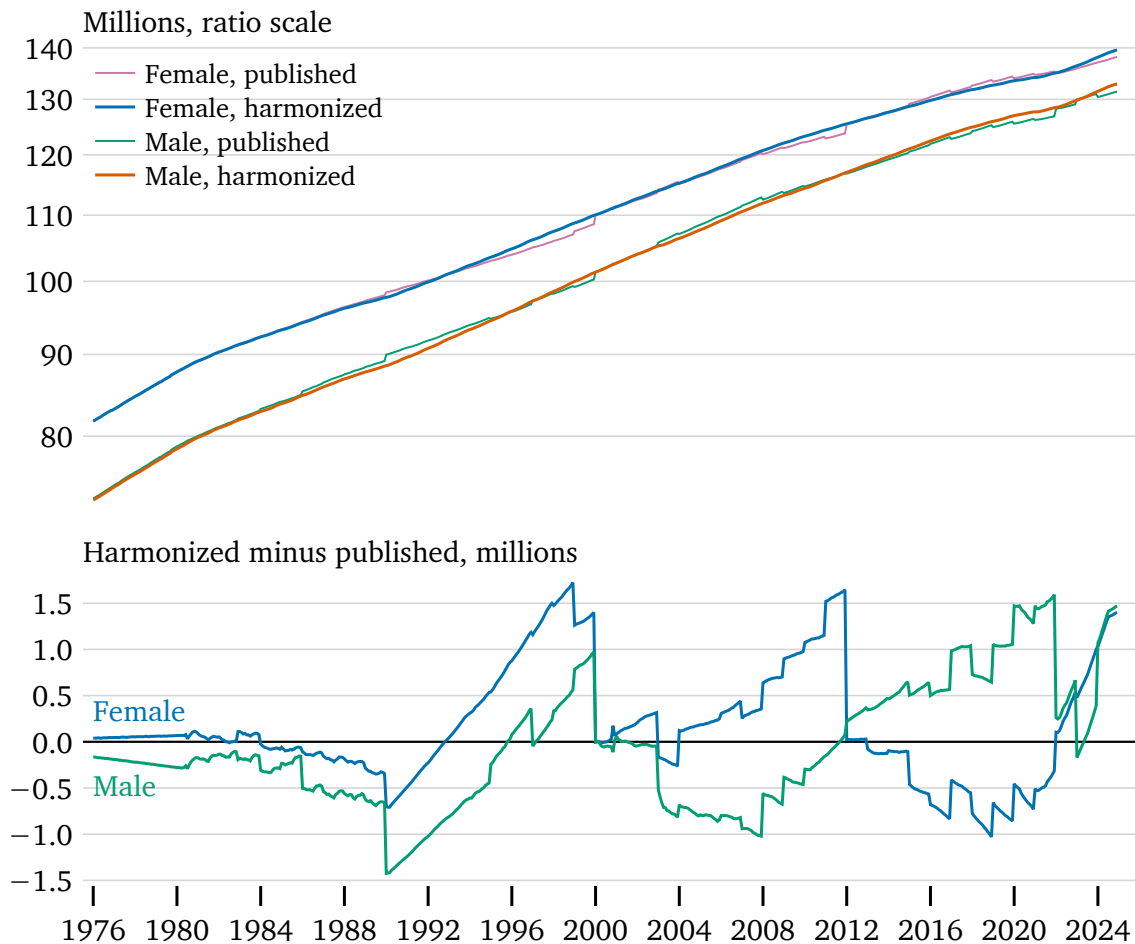
As shown in figure 5, our harmonized estimate of the LFPR (the blue line) differs from the published series (the orange line).<sup>15</sup> Our harmonized estimate is more than 0.4 percentage point (pp) above the official estimate in 1976, mostly reflecting the adjustment for the CPS redesign, which boosts our estimate by 0.4 pp over 1976–93. (The upward shift in the published LFPR in January 1994 is visible in the upper panel.) The published series jumps above our estimate in 1994, because we adjust for the CPS redesign while the published series does not, and remains about 0.2 pp above over the 2000s. The published series falls below our harmonized series in 2012, when the BLS first incorporated population estimates based on the 2010 Census into the CPS, with the gap rising to about 0.4 pp at the start of 2020. The published LFPR jumped more than 0.2 pp in January 2022, when population estimates based on the 2020 Census were introduced, narrowing much of the gap to the harmonized series. It is notable that the harmonized

---

14. In the past, the BLS has produced research series of various labor force statistics that seek to smooth some breaks in the labor force statistics that result from the switching to new population estimates each January (DiNatale 2003). More recently, the BLS produced experimental series for the level of the labor force, employment, and unemployment from April 2020 forward that seek to account for the sizable revisions to the population in the V2024 estimates (Bureau of Labor Statistics 2025). Section E in the appendix discusses the differences between our approach and the BLS's experimental series.

15. Figure D5 in the appendix shows the implications for men's and women's LFPR.

**Figure 4. Civilian noninstitutionalized population, harmonized and published**



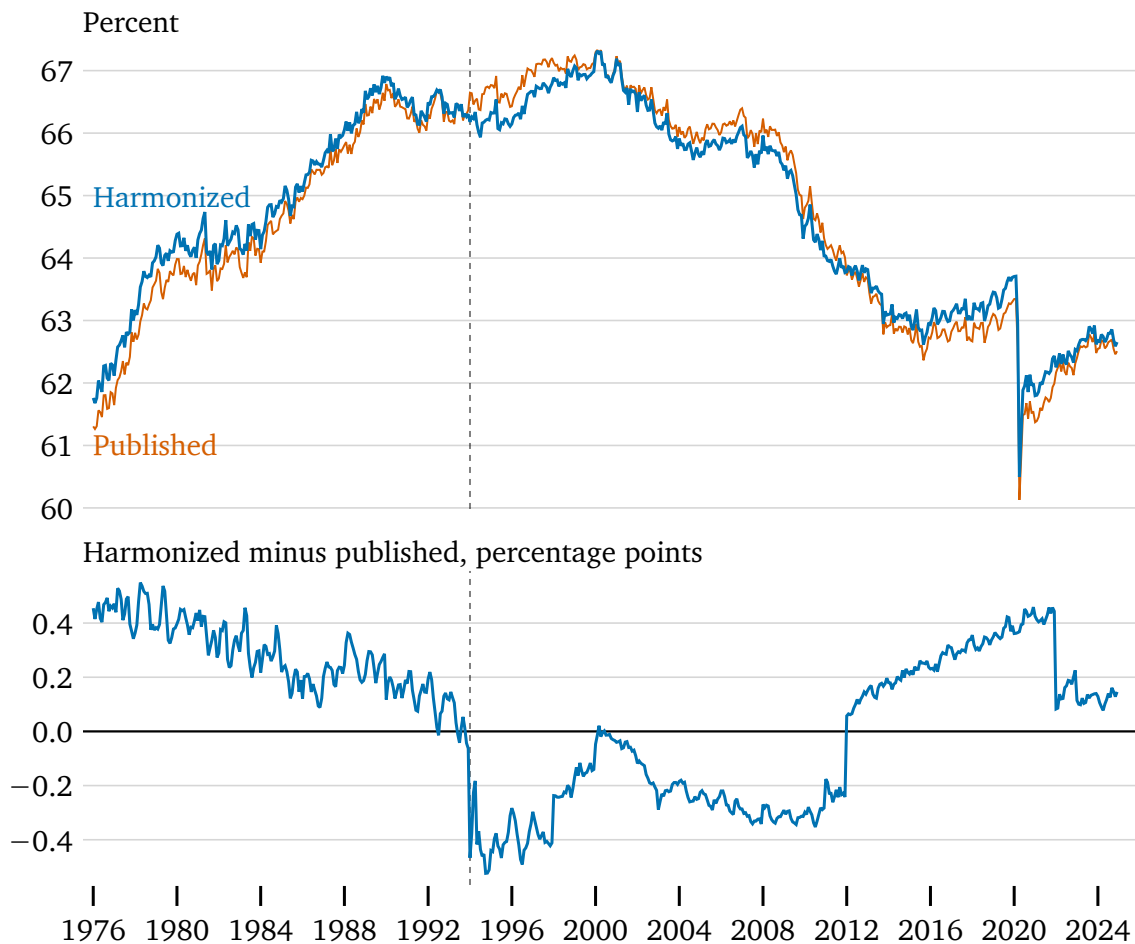
Note: Individuals aged 16 or older.

Source: Bureau of Labor Statistics (BLS); authors' calculations using data from the Census Bureau and the BLS.

LFPR is both above and below the published LFPR at various points over time, with the difference between the two changing sign several times.

Despite the sizable differences in population and the LFPR, our estimate for the unemployment rate is generally quite close to the published series. In fact, the two series in figure 6 are well within one-tenth from 1994 forward. The similarity of these two series likely results from the unemployment rate varying less across demographic groups than does the LFPR, so revisions to population shares (especially for older individuals) will have a larger effect on the LFPR relative to their effect on the unemployment rate. Before 1994, the harmonized series is somewhat above the published series, reflecting two factors. First, adjusting for the CPS redesign added 0.1 pp to

**Figure 5. Labor force participation rate, harmonized and published**



Note: Individuals aged 16 or older. Vertical line denotes January 1994, when the Current Population Survey redesign was implemented.

Source: Bureau of Labor Statistics (BLS); authors' calculations using data from the Census Bureau and the BLS.

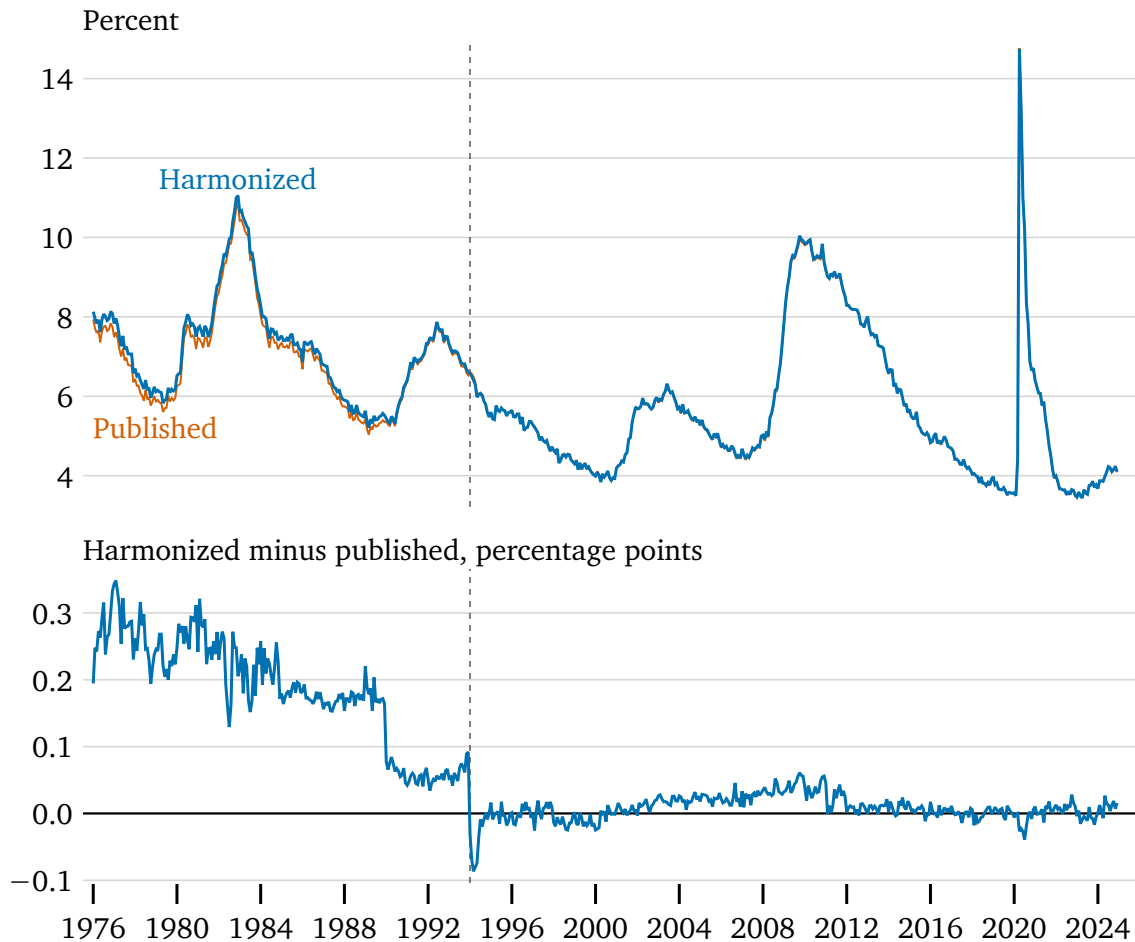
the pre-1994 series. Second, the introduction of population estimates based on the 1990 Census raised the relative population and labor force shares of younger people, who tend to have higher-than-average unemployment rates.<sup>16</sup>

## 5.2. Recovery of labor force participation after the COVID-19 pandemic

Accounting for population revisions has been important for understanding the recovery in the LFPR following the pandemic. Updated population es-

16. Figure D6 in the appendix shows the joint implications of the LFPR and unemployment rate for the employment-to-population ratio ratio.

**Figure 6. Unemployment rate, harmonized and published**

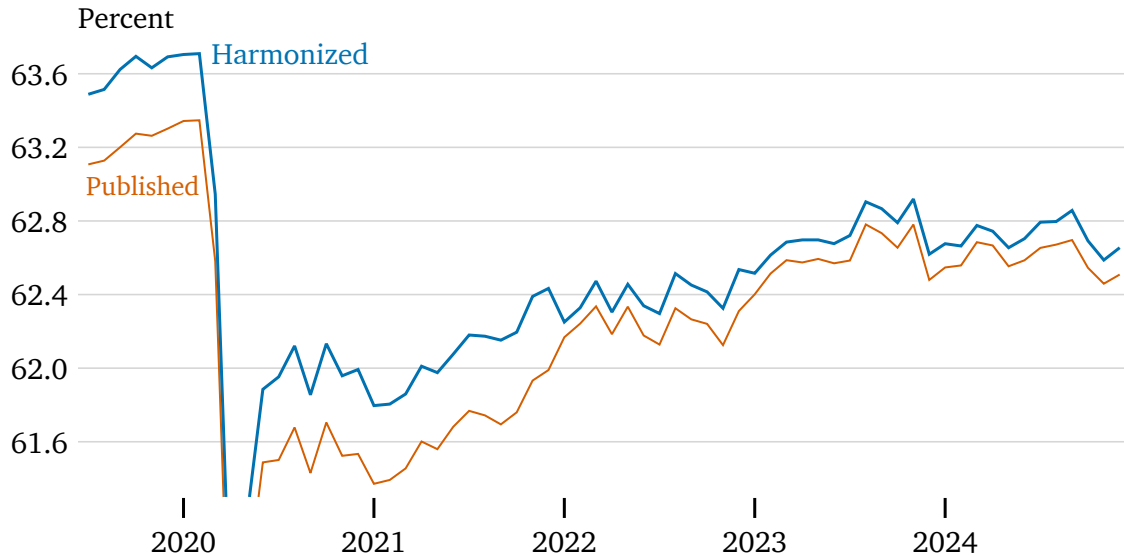


Note: Individuals aged 16 or older. Vertical line denotes January 1994, when the Current Population Survey redesign was implemented.

Source: Bureau of Labor Statistics (BLS); authors' calculations using data from the Census Bureau and the BLS.

timates released since 2020 have seen fairly large revisions from several novel factors, including the incorporation of the 2020 Census and revisions to mortality source data to reflect excess deaths during the pandemic and updated estimates of immigration capturing the 2021–24 surge in net entry. Each of these factors is likely to change population differentially across demographic groups, especially across the age distribution, which has important consequences for the LFPR. Our approach corrects for these factors consistently across the history of the series, allowing for separate revisions to each period depending on the underlying source data.

**Figure 7. Labor force participation during and after the pandemic**



Source: Bureau of Labor Statistics (BLS); authors' calculations using data from the Census Bureau and the BLS.

Figure 7 compares our LFPR estimate with the published series, showing several notable differences. First, our estimate is 0.4 pp higher than the published series on the eve of the pandemic in February 2020. This difference stems largely from the difference between the PEP estimates over the 2010s and the final count from the April 2020 decennial census, the latter of which saw upward revisions to population for high-LFPR age groups, particularly men aged 35 to 60, and downward revisions to population for low-LFPR age groups, especially men and women over 75. Incorporating this revision into the population estimates, which the published data do not, results in a higher estimate for the LFPR pre-pandemic.

Second, our estimate shows a more consistent pace of recovery from mid-2020 through mid-2023. Over this period, the LFPR rose  $\frac{1}{4}$  pp per year, with only modest fluctuations around this trend. In contrast, the published data show a notable pickup in the pace of recovery of the LFPR in late 2021 into early 2022. However, this difference in the pace of the LFPR recovery is entirely attributable to the population revision implemented in January 2022, which revised up the LFPR by 0.2 pp that month. Smoothing out this break, as our estimate does, reveals the steady pace of the LFPR improvement in the post-pandemic period.

Third, as a result of these differences, our estimate delivers a more accurate assessment of the gap between the pre- and post-pandemic LFPR. The population revision implemented in January 2022 incorporated both the new decennial census data for 2020 and updated estimates of mortality for 2020. Each of these factors could explain why the population of low-LFPR age groups revised down, and thus why the LFPR in early 2022 revised up, but they would have very different implications for whether the February 2020 LFPR would revise up. Our approach relies on the underlying population estimates to distinguish between these factors, so we do not need to make assumptions about how historical population revised. In this case, our estimate for the February 2020 LFPR also revised up by 0.2 pp, indicating that the 2020 Census data were the dominant factor. As a result, the gap between the pre- and post-pandemic LFPR in our estimates was unchanged from this revision (since both revised up 0.2 pp), while the gap for published estimates revised to be 0.2 pp narrower. In this way, the published estimates overestimated the LFPR recovery relative to pre-pandemic conditions starting in 2022, while our estimates do not.

### 5.3. Trends in native-born population

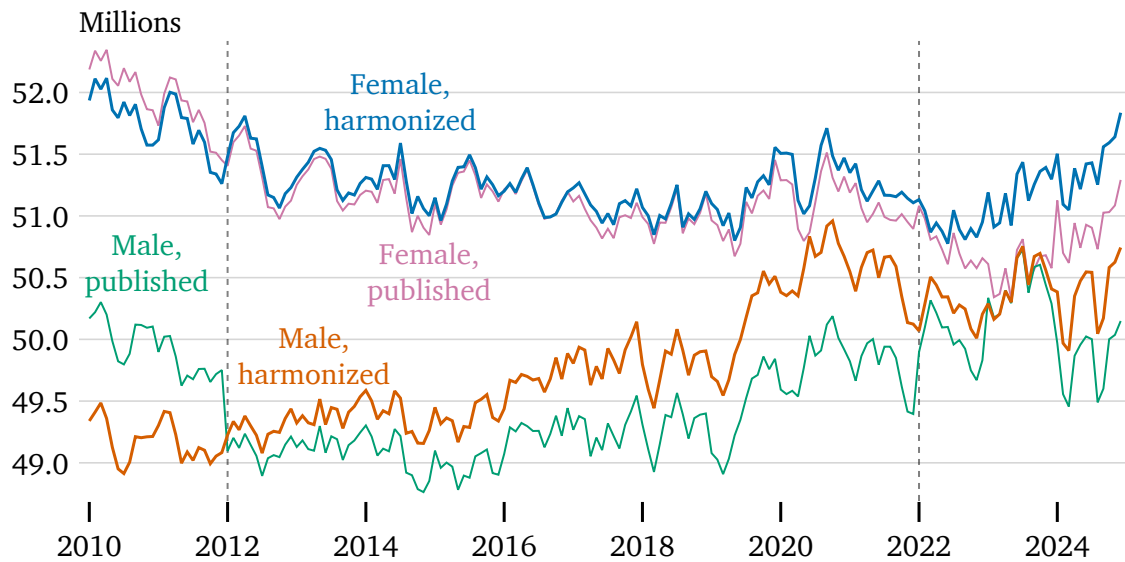
An additional example of the value of our approach comes from examining trends in the native-born population in recent years.<sup>17</sup> Note that nativity is not a targeted characteristic, although it is likely to be correlated with the intersection of demographic characteristics that are targeted. Figure 8 shows estimates of the prime-age (25 to 54) native-born population for men (in green) and women (in pink), where these series are calculated from the published CPS weights. The series for men *declines* over 2010–15 (with a particularly sharp drop in January 2012), a period when the prime-age population should be increasing. In addition, the population for women declines over 2020–23, while the population for men moves sideways on net. These patterns are puzzling and make taking these estimates at face value difficult.

However, the unusual pattern when using published microdata weights stems from the inconsistencies introduced by outdated population estimates. The blue and orange lines in figure 8 show estimates of the same

---

17. We thank Brian Kovak for bringing this example to our attention.

**Figure 8. Native-born prime-age population for men and women**



Note: Vertical lines denote the introduction of population controls drawn from a new decennial census (January of 2012 and 2022).

Source: Bureau of Labor Statistics (BLS); authors' calculations using data from the Census Bureau and the BLS.

series using our weights constructed to be consistent with our harmonized population estimates. Using our harmonized estimates, we find that the native-born prime-age male population rose throughout the 2010s and was relatively flat since then—a very different pattern than when using the published weights. These series show similar patterns for men and women following the pandemic, with a decrease over 2020–23 and then flattening afterward. Correcting for population revisions makes it clear that the pattern for women is not an anomaly; the same forces are leading to similar outcomes for men, but this fact is obscured in the published data due to different revisions to the population of men and women.

This example highlights a major advantage of our approach — namely, that it can produce a corrected estimate of *any* series calculable from the CPS microdata. Our approach not only delivers new estimates of the major headline series that the BLS releases, but it also allows for calculating series for more detailed demographic groups than the BLS releases while maintaining the consistency with harmonized population estimates. Researchers wanting to account for updated population estimates in their analysis do not need to use a separate data source or apply any form of raking — they

can simply repeat their analysis using our updated weights and get new estimates.

#### 5.4. Immigration adjustment for 2021 to present

In January 2024, the CBO released its annual demographic projections (Congressional Budget Office 2024). This usually staid report was a blockbuster, as the CBO estimated a substantially higher level of population than official estimates from the Census Bureau indicated at the time, with higher immigration accounting for nearly all of the difference. This upward revision largely stemmed from information in administrative records tabulated by the Department of Homeland Security (DHS) that measured direct flows across the border, which indicated much higher rates of inflows in recent years than previous methods captured. With the Census Bureau's next vintage of population estimates, the V2024 estimates released in December 2024, the Census Bureau updated its methodology to incorporate administrative data on immigration inflows, similar to those used by the CBO.<sup>18</sup> As a result, the official V2024 population estimate for December 2024 (the blue line in figure 9) revised up by 3.5 million compared with the Census Bureau's previous estimate (the orange line).

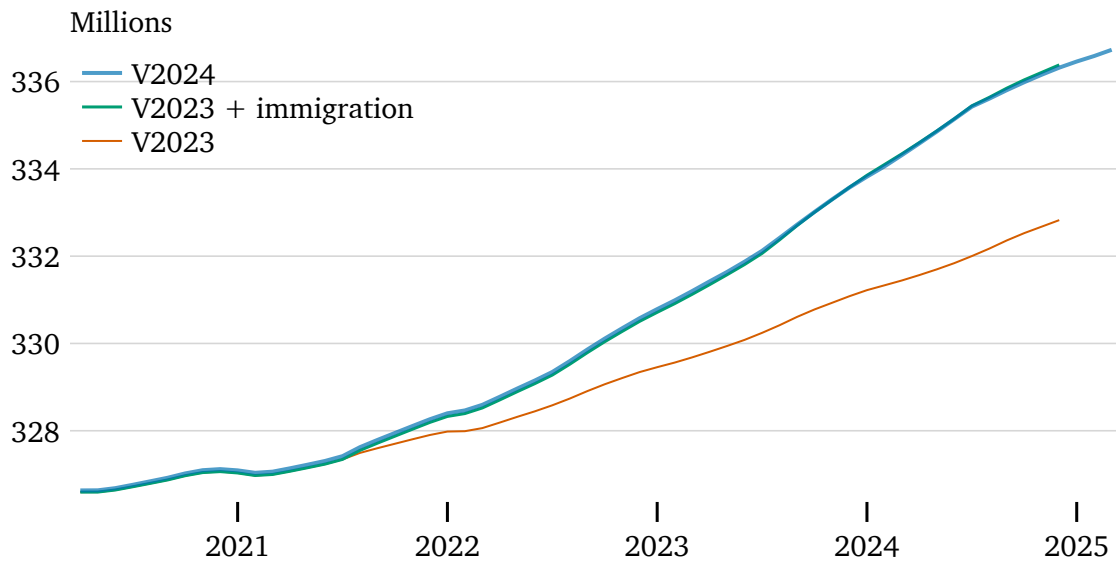
An advantage of our method is that it can be used to estimate microdata weights and labor force statistics consistent with alternative population estimates, as long as one can estimate the population at a sufficiently detailed level. This section summarizes how our method could have been used in advance of the V2024 estimates to update microdata weights and labor force statistics to account for higher immigration. Section F of the appendix provides a more detailed description.

Nearly all the revision to the Census Bureau's population estimates from V2023 to V2024 can be attributed to revisions in estimated net immigration. Thus, we can achieve a close approximation of the V2024 estimates if we assume that the V2023 population estimates are accurate for the native-born population and solely focus on estimating the revision to net immigration from V2023 to V2024.

---

18. Gross et al. (2024) summarizes the V2024 estimates and sources of revision.

**Figure 9. Recent vintages of the civilian noninstitutionalized population**



Note: Series labeled “V2023 + immigration” adds to the V2023 series our estimate of immigration inflows; both series extend through December 2024 and overlap for most of the period.

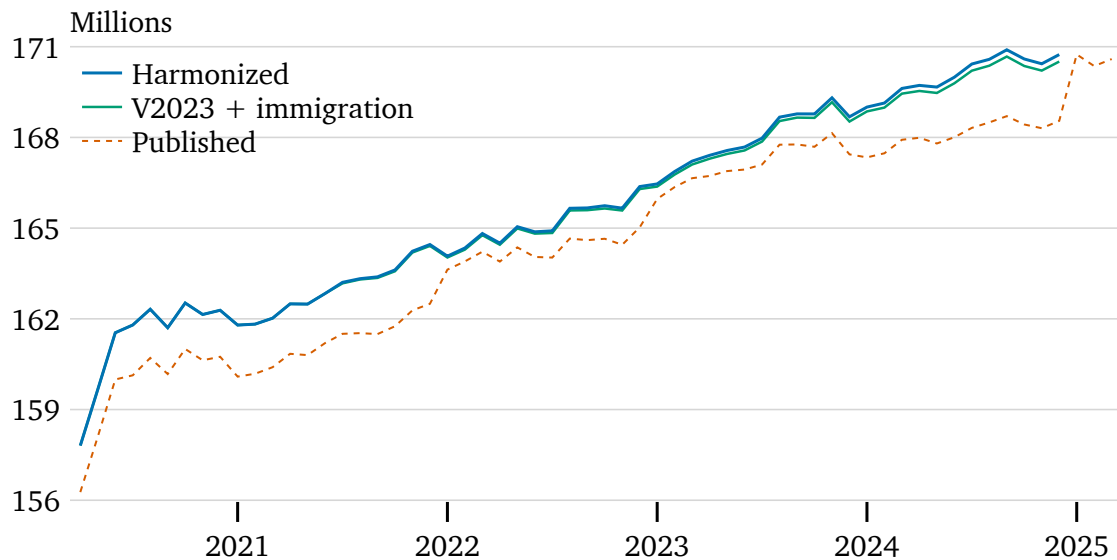
Source: Census Bureau; authors’ calculations.

The primary source of the revision to the Census Bureau’s estimate of net immigration was incorporating administrative data sources on humanitarian migration from July 2021 through June 2024.<sup>19</sup> Our calculations use the same monthly administrative data on humanitarian migration as those used by the Census Bureau.<sup>20</sup> Because a portion of these inflows were likely already captured by the American Community Survey (ACS) data, we follow the Census Bureau’s approach of assuming that only 75 percent of the humanitarian migration inflows should be passed through to the revision to net immigration. After this scaling, humanitarian migration boosts the average pace of net immigration by about 75,000 per month over those three years.

19. Additionally, part of the revision likely stems from updating the ACS source data. For instance, the V2024 net immigration totals use the 2023 ACS to estimate unadjusted foreign-born immigration for July 2022 to December 2024, whereas the V2023 estimates had to rely on the 2022 ACS. We calculate that switching from the 2022 to the 2023 ACS data boosted the V2024 estimate of net immigration by about 20,000 per month for this 30-month period.

20. We use administrative data from the DHS on monthly inflows of migrants who were either released by U.S. Border Patrol with a notice to appear before an immigration judge or granted parole by the Office of Field Operations. We also use data published by the Department of State’s Refugee Processing Center on the monthly number of refugee admissions.

**Figure 10. Labor force level**



Note: The harmonized and “V2023 + immigration” series extend through December 2024 and overlap for most of the period.

Source: Bureau of Labor Statistics (BLS); authors’ calculations.

Next, we distribute the revision to foreign-born immigration to the demographic groups that we control our harmonized CPS estimates to. Specifically, we sum humanitarian migration by region of origin and year and distribute those totals to characteristics using proxy universes of recently arrived foreign-born individuals tabulated from the ACS. We then sum across regions to get immigration-adjusted population estimates.

By mapping the revision to net immigration into the revision for detailed demographic groups, we can approximate the Census Bureau’s V2024 headline population estimates. Returning to figure 9, we see that our immigration-adjusted estimates (the green line) are nearly indistinguishable from the V2024 estimates (the blue line).

We next calculate new microdata weights that control to this alternative population and, using those weights, time-series estimates of the labor force statistics. As shown in figure 10, the labor force level estimated from our immigration-adjusted population (the green line) is very close to the harmonized series (the blue line), which is based on the V2024 population estimates. Indeed, our alternative estimates correctly identify a modest positive effect on the LFPR (0.12 pp, figure 11a) and a negligible effect on the unemployment rate (0.02 pp, figure 11b) as of December 2024. These results

demonstrate the advantages of being able to assess the effect of alternative population estimates in advance of revisions to the official population estimates.

## 6. Conclusion

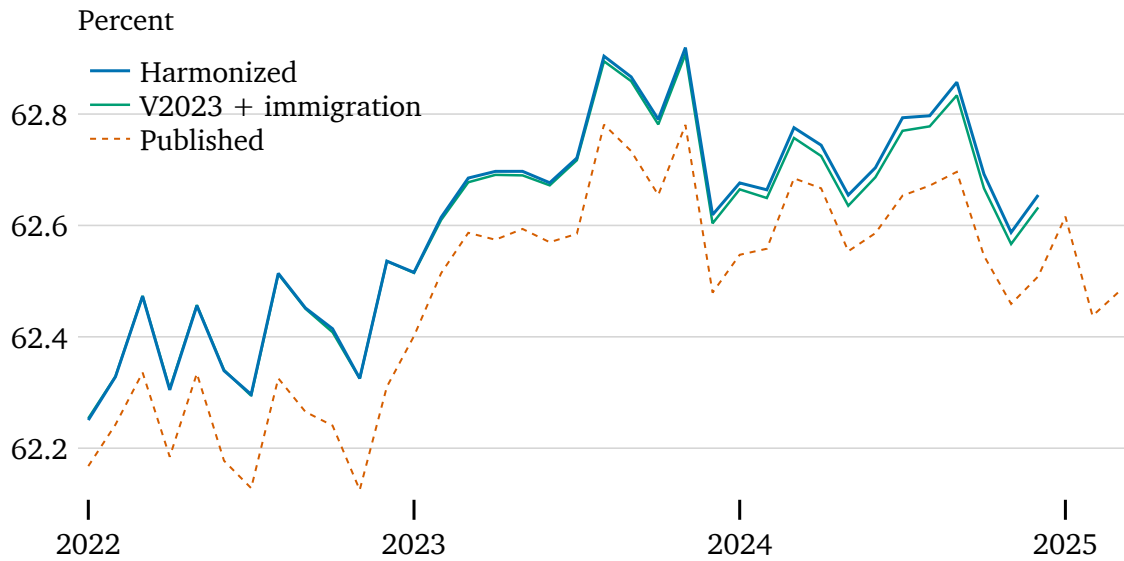
In this paper, we outline a methodology for producing harmonized population and labor force statistics. Our approach addresses the issue of yearly discontinuities in official statistics and CPS microdata, providing a set of time series consistent with the latest population estimates. These series shed new light on the post-pandemic labor market recovery, indicating that the labor force shortfall in recent years has been about 1.5 million larger than published data indicate. We provide harmonized time series of the population and key labor force statistics extending back to 1948 for all major demographic groups. In addition, the accompanying microdata weights also offer researchers a straightforward way to reproduce any CPS calculation or statistic adjusted for population revisions. We plan to update both the microdata weights and the harmonized time series annually to reflect each new vintage of population estimates from the Census Bureau.

Although we focus on harmonizing labor force statistics to reflect the latest official population estimates, our approach can also be used to produce labor force statistics consistent with alternative population estimates. Our estimates incorporate the latest Census Bureau population estimates, which account for the upward surprise to net immigration into the United States in recent years. However, if immigration continues to outpace projections (or falls below current projections), one would only need to add an estimate of this error to (or subtract an estimate of it from) the demographic-specific population estimates and then follow our methodology to produce time series that reflect these population surprises.

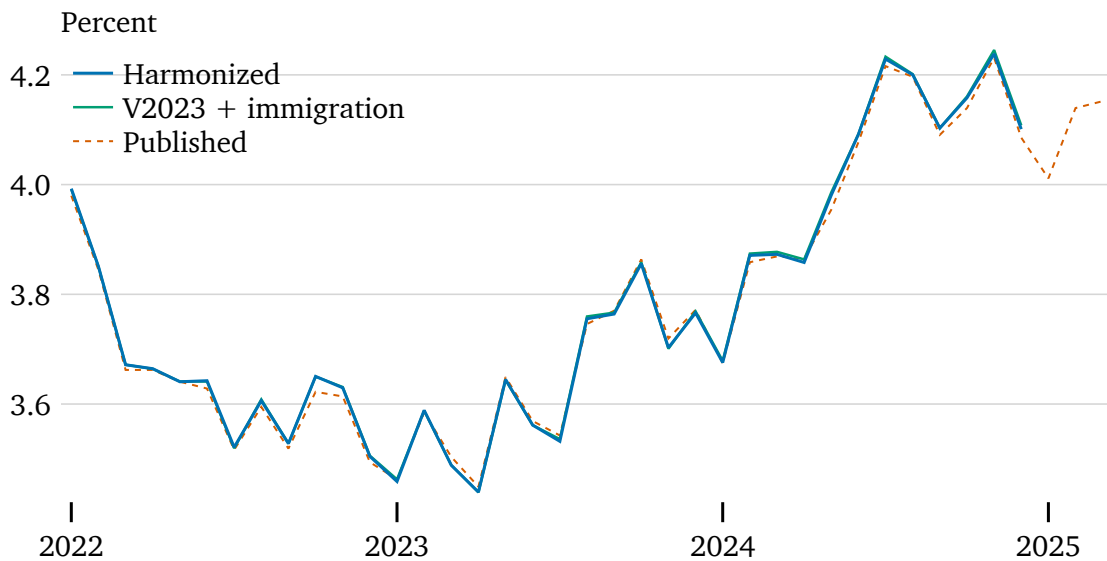
There are additional improvements to our methodology that remain for future work. We have replicated the BLS's procedures for weighting to match demographic population controls, but in the future, it would be helpful to additionally replicate matching the geographic targets as well. Our approach makes several simplifying assumptions, including linearly distributing the decennial census surprises over the preceding decade, using civil-

**Figure 11. Effects of alternative population estimates on labor force statistics**

**(a) LFPR**



**(b) Unemployment rate**



Note: The harmonized and “V2023 + immigration” series extend through December 2024 and overlap for most of the period.

Source: Bureau of Labor Statistics (BLS); authors’ calculations.

ian population estimates to impute the CNP for 1980–90, and distributing multiple-race-identifying respondents to harmonize the change in race categories, each of which could be oversimplifying and may merit further adjustments. We also have not explored estimating seasonal factors for our harmonized series, which are likely similar to the seasonal factors for published series that we use, but could differ somewhat in practice.

## 7. References

- Bezanson, Jeff et al. (2017). “Julia: A Fresh Approach to Numerical Computing”. In: *SIAM Review* 59.1, pages 65–98. DOI: [10.1137/141000671](https://doi.org/10.1137/141000671). URL: <https://epubs.siam.org/doi/10.1137/141000671>.
- Bouchet-Valat, Milan and Bogumił Kamiński (2023). “DataFrames.jl: Flexible and Fast Tabular Data in Julia”. In: *Journal of Statistical Software* 107.4, pages 1–32. DOI: [10.18637/jss.v107.i04](https://doi.org/10.18637/jss.v107.i04). URL: <https://www.jstatsoft.org/index.php/jss/article/view/v107i04>.
- Breau, Patricia and L Ernst (1983). “Alternative estimators to the current composite estimator”. In: *Proceedings of the Section on Survey Research Methods, American Statistical Association*. Volume 397402.
- Bureau of Labor Statistics (May 1962). “May 1962”. In: *Employment and Earnings* 8.11. URL: <https://fraser.stlouisfed.org/title/60#20189>.
- Bureau of Labor Statistics (Feb. 1972). “February 1972”. In: *Employment and Earnings* 18.8. URL: <https://fraser.stlouisfed.org/title/60#20279>.
- Bureau of Labor Statistics (Feb. 1974). “February 1974”. In: *Employment and Earnings* 20.8. URL: <https://fraser.stlouisfed.org/title/60#20303>.
- Bureau of Labor Statistics (Feb. 2006). “February 2006”. In: *Employment and Earnings* 53.2. URL: <https://fraser.stlouisfed.org/title/60#19934>.
- Bureau of Labor Statistics (Feb. 2025). *Experimental Series Accounting for January 2025 Population Control Effects*. URL: <https://www.bls.gov/cps/methods/population-controls/experimental-series-accounting-for-January-2025-population-control-effects.htm>.
- Congressional Budget Office (2024). *The Demographic Outlook: 2024 to 2054*. URL: <https://www.cbo.gov/publication/59899>.

- Cooper, Daniel H. et al. (2021). “Population Aging and the US Labor Force Participation Rate”. In: *Current Policy Perspectives* 93533.
- Danisch, Simon and Julius Krumbiegel (2021). “Makie.jl: Flexible high-performance data visualization for Julia”. In: *Journal of Open Source Software* 6.65, page 3349. DOI: [10.21105/joss.03349](https://doi.org/10.21105/joss.03349). URL: <https://doi.org/10.21105/joss.03349>.
- Department of Economic and Social Affairs, Statistics Division (1999). *Standard Country or Area Codes for Statistical Use*. 49. United Nations. URL: [https://unstats.un.org/unsd/publication/SeriesM/Series\\_M49\\_Rev4\(1999\)\\_en.pdf](https://unstats.un.org/unsd/publication/SeriesM/Series_M49_Rev4(1999)_en.pdf).
- Department of Homeland Security, Office of Homeland Security Statistics (2025). *Immigration Enforcement and Legal Processes Monthly Tables through November 2024*. Accessed: 2025-05-22. URL: <https://ohss.dhs.gov/topics/immigration/immigration-enforcement/monthly-tables>.
- Department of State, Refugee Processing Center (2025). *Refugee Admissions Report as of December 31, 2024*. Accessed: 2025-05-22. URL: <https://www.wrapsnet.org/documents/PRM%20Refugee%20Admissions%20Report%20as%20of%2031%20Dec%202024.xlsx>.
- DiNatale, Marisa L. (2003). *Creating Comparability in CPS Employment Series*. URL: <https://www.bls.gov/cps/methods/population-controls/cpscomp.pdf>.
- Drew, Julia A. Rivera, Sarah Flood, and John Robert Warren (2014). “Making Full Use of the Longitudinal Design of the Current Population Survey: Methods for Linking Records across 16 Months”. In: *Journal of Economic and Social Measurement* 39, pages 121–144. DOI: [10.3233/JEM-140388](https://doi.org/10.3233/JEM-140388).
- Edelberg, Wendy and Tara Watson (Mar. 2024). *New Immigration Estimates Help Make Sense of the Pace of Employment*. Working paper. The Hamilton Project.
- Flood, Sarah et al. (2024). *IPUMS CPS: Version 12.0 [dataset]*. DOI: [10.18128/D030.V12.0](https://doi.org/10.18128/D030.V12.0).
- Forsythe, Eliza et al. (2022). “Where have all the workers gone? Recalls, retirements, and reallocation in the COVID recovery”. In: *Labour Economics* 78, page 102251.
- Gross, Mark et al. (Dec. 2024). *Census Bureau Improves Methodology to Better Estimate Increase in Net International Migration*. URL: <https://www.census.gov/newsroom/blogs/random-samplings/2024/12/international-migration-population-estimates.html>.

- Hobijn, Bart and Ayşegül Şahin (2022). “‘Missing’ Workers and ‘Missing’ Jobs Since the Pandemic”. In: *FRB of Chicago Working Paper* 2022-54.
- Huang, Elizabeth T. and Lawrence R. Ernst (1981). “Comparison of an alternative estimator to the current composite estimator in CPS”. In: *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pages 303–308.
- Kolko, Jed (Apr. 2025). *Historically Comparable CPS Microdata Weights*. URL: <https://jedkolko.com/cps-weights/>.
- Madrian, Brigitte C. and Lars John Lefgren (2000). “An Approach to Longitudinally Matching Current Population Survey (CPS) Respondents”. In: *Journal of Economic and Social Measurement* 26.1, pages 31–62.
- Office of Management and Budget (Oct. 1997). “Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity”. In: *Federal Register* 62.210. URL: <https://www.govinfo.gov/content/pkg/FR-1997-10-30/pdf/97-28653.pdf>.
- Polivka, Anne E. and Stephen M. Miller (1998). “The CPS after the Re-design: Refocusing the Economic Lens”. In: *Labor Statistics Measurement Issues*. University of Chicago Press, pages 249–89.
- Robertson, John (Feb. 2023). *Population Control Adjustment’s Impact on Labor Force Data: The 2023 Edition*. URL: <https://www.atlantafed.org/blogs/macroblog/2023/02/09/population-control-adjustments-impact-on-labor-force-data--2023-edition>.
- Robertson, John and Jon Willis (Mar. 2022). *Assessing Recent Labor Market Improvement*. URL: <https://www.atlantafed.org/blogs/macroblog/2022/03/01/assessing-recent-labor-market-improvement>.
- Schmitt, John (Aug. 2003). *Creating a Consistent Hourly Wage Series from the Current Population Survey’s Outgoing Rotation Group, 1979–2002*. Working paper. Center for Economic and Policy Research.
- Serrato, Juan Carlos Suárez and Philippe Wingender (2016). *Estimating Local Fiscal Multipliers*. Working paper 22425. National Bureau of Economic Research.
- Stephan, Frederick F. (1942). “An Iterative Method of Adjusting Sample Frequency Tables when Expected Marginal Totals Are Known”. In: *The Annals of Mathematical Statistics* 13.2, pages 166–178.
- U.S. Census Bureau (Mar. 2002). *Current Population Survey Design and Methodology*. Technical Paper 63RV.
- U.S. Census Bureau (Oct. 2006). *Current Population Survey Design and Methodology*. Technical Paper 66.

U.S. Census Bureau (Oct. 2019). *Current Population Survey Design and Methodology*. Technical Paper 77.

U.S. Census Bureau (Nov. 2024a). *Methodology for the Intercensal Population and Housing Unit Estimates: 2010 to 2020*. URL: <https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/intercensal/2010-2020-intercensal-estimates-methodology.pdf>.

U.S. Census Bureau (Dec. 2024b). *Methodology for the United States Population Estimates: Vintage 2024*. Technical report. URL: <https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2020-2023/methods-statement-v2024.pdf>.

## Appendix

The data work and calculations in this paper are performed using the [Julia programming language](#) (Bezanson et al. 2017). Tabular data manipulation comes from the [DataFrames](#) package (Bouchet-Valat and Kamiński 2023). Graphics are created using the [Makie](#) plotting library (Danisch and Krumbiegel 2021).

### A. Harmonized population estimates

We construct a harmonized monthly time series for the civilian noninstitutionalized population (CNP) by age, sex, race, and Hispanic origin from January 1976 through December 2024. This requires merging population estimates drawn from 6 different decennial censuses:

- 1970 Census base (January 1976–March 1980)
- 1980 Census base (April 1980–March 1990)
- 1990 Census base (April 1990–December 2002)
- 2000 Census base (January 2003–March 2010)
- 2010 Census base (April 2010–March 2020)
- 2020 Census base (April 2020–present)

Our harmonized estimates are derived primarily from data published by the Census Bureau’s [Population Estimates Program \(PEP\)](#).<sup>21</sup> The Census Bureau produces two types of population estimates:

---

21. This section summarizes information from U.S. Census Bureau (2024a,b).

1. *Postcensal estimates* are time series of the population since the most recent decennial census, constructed using measures of population change. In particular, the population estimate at a given date starts from a population base (e.g., the previous decennial census or the previous date in the time series) and adds births, subtracts deaths, and adds net migration. Each vintage of estimates includes all years since the most recent decennial census and supersedes any previously produced estimates, as the latest vintage is based on more up-to-date data.
2. *Intercensal estimates* smooth the transition from one decennial census count to the next. Specifically, they reconcile the final postcensal estimates from a census base with the actual count from the subsequent census in order to provide a consistent time series of population estimates over the decade that reflects the latest census results.

Thus, the intercensal estimate represents the best estimate of the monthly path of the population over a decade that is consistent with both the starting and ending census bases, which are the most accurate counts of the population at those two points in time. For example, the latest intercensal estimates adjust the final 2010–20 postcensal estimates — which start from April 1, 2010 and extend through April 1, 2020 — to account for differences between the postcensal estimates for April 1, 2020 and the 2020 Census. (See figure 3 in the main text for an example.)

Technically, because the relevant population for the Current Population Survey (CPS) is the CNP, we use the CNP “estimates base” as the starting point for each vintage of population estimates rather than the “census base”. In addition, the PEP’s CNP estimates based on the 2020 Census use a “blended base” — which uses the full count for aggregate statistics and estimates to get demographic subtotals — as the starting point (U.S. Census Bureau 2024b).

#### A1. Calculating intercensal estimates

The official intercensal estimates by characteristics are produced only for the resident population or only at annual frequency. Thus, we construct our own monthly estimates for the CNP using the Census Bureau’s latest methodology (U.S. Census Bureau 2024a). For most periods, the PEP pub-

lishes estimates of the monthly national CNP by age, sex, race, and Hispanic origin. (We refer to age, sex, race, and Hispanic origin collectively as “characteristics”.)

We start from the PEP’s final postcensal estimates for a decade and calculate the difference between the April 1st population estimate (which is based on the previous census) and the current census population estimates base. This difference is then distributed linearly over the decade:

$$(A.1) \quad P_{it} = Q_{it} + t \times \left( \frac{P_{iD} - Q_{iD}}{t_D} \right),$$

where

$t$  = Days since April 1 of previous census base

$D$  = April 1 of the current census base (3652 or 3653, depending on the number of leap years)

$P_{it}$  = Intercensal estimate for group  $i$  at day  $t$

$Q_{it}$  = Postcensal estimate for group  $i$  at day  $t$

$P_{iD}$  = April 1 population estimates base for group  $i$  from current census

$Q_{iD}$  = April 1 postcensal estimate for group  $i$  based on previous census.

We estimate these smoothed time series separately for each group (defined by single-year age, sex, race, and Hispanic origin). One potential downside of linear interpolation is that it can result in negative values, particularly when the time-series values are small and the variation over time is large relative to the mean. To avoid these situations, we aggregate ages 90 or older for a given sex, race, and Hispanic origin before constructing our intercensal estimates. This is sufficient for equation (A.1) to be weakly positive for all groups.

The remaining subsections provide details about each intercensal estimate.

## A2. 1970 Census base (January 1976 to March 1980)

The Census Bureau does not have monthly estimates of CNP by characteristics before the 1980 Census base, so we used the Bureau of Labor Statistics (BLS)’s published estimates of the CNP by sex, age group, and race as inde-

pendent targets. These characteristic groups become the starting points for calculating the population targets.

We take the monthly published population for 48 characteristic groups by sex, age (16 to 17, 18 to 19, 20 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 and older), and race (White, Black, and all other races). Although there are no breaks in the published levels — the series are post-censal estimates from the 1970 Census base — we need to construct an intercensal estimate so that they match the 1980 Census base. To do this, we calculate the April 1980 CNP for these groups from the 1980 Census base (described in the next subsection) and smooth the error of closure back to April 1970.

#### **A3. 1980 Census base (April 1980 to March 1990)**

There are no monthly population estimates by characteristics. We start with quarterly estimates of the civilian population (CP) by characteristics and interpolate them to monthly. We then sum the interpolated monthly values to a national total and rake that to the published monthly national population total.

Next we create an estimate of the CNP. From the 1990 Census base, which reported estimates by characteristics for the CP and CNP, we calculate the ratio of the CNP to CP on April 1, 1990. We then multiply our CP estimates from the 1980 Census base by that ratio.

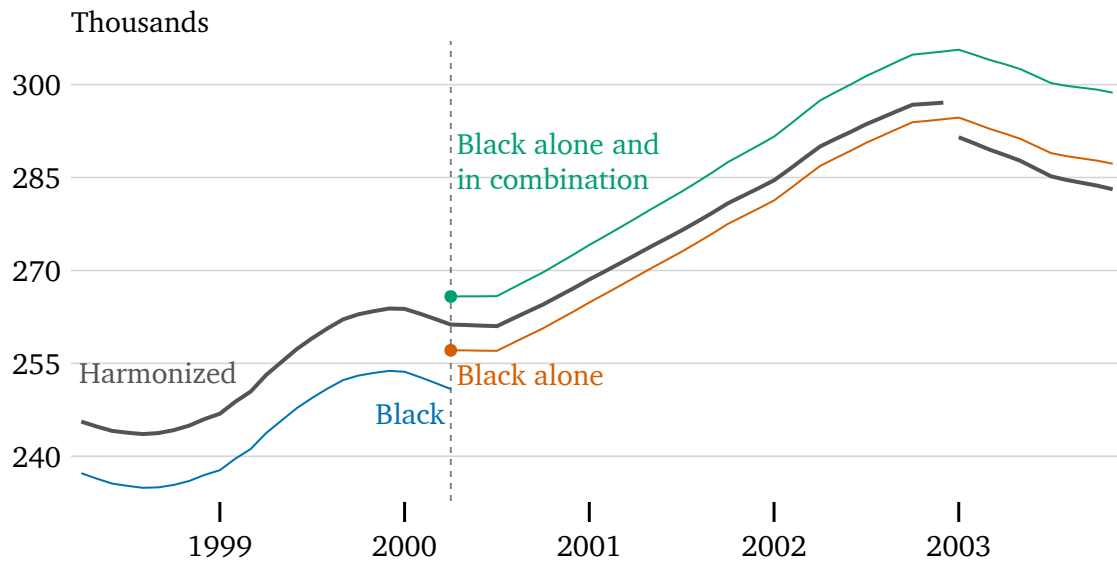
#### **A4. 1990 Census base (April 1990 to December 2002)**

As described in Office of Management and Budget (1997), two major changes to the race categories took place between the 1990 and 2000 censuses:

1. In previous censuses, responses to the race question were limited to a single category; in 2000, for the first time, respondents could check as many boxes as necessary to identify their race.
2. The “Asian and Pacific Islander” category was separated into “Asian” and “Native Hawaiian or Other Pacific Islander”.

Before January 2003, the CPS was controlled to 3 race categories: “White”, “Black”, and “all other races”. (Hispanic origin is considered an ethnicity, not a race. Hispanics may be of any race.) From January 2003 forward,

**Figure A1. Population of 22-year-old non-Hispanic Black women**



Note: Solid circles denote the 2000 Census base. The break in the harmonized series in January 2003 reflects a change in the concept of race.

Source: Authors' calculations using data from the Census Bureau.

the CPS was controlled to 4 race categories in the national coverage step (“White alone”, “Black alone”, “Asian alone”, and “all other races”) and to 3 race categories in the second-stage race-sex-age step (“White alone”, “Black alone”, and “all other races”).

Figure A1 shows an example of how we harmonized the population estimates by race. The blue line is the population for 22-year-old non-Hispanic “Black” women from the Vintage 2000 postcensal estimate. Before the 2000 Census, the category “Black” included some individuals who would have selected multiple races if they had the choice to do so. From April 2000 forward, the Census Bureau’s estimates are reported for “Black alone” (the orange line) and “Black alone and in combination” (the green line). Our harmonized estimate (the black line) converts the estimates from a multi-race basis to a single-race basis over 2000–02. The new race categories were first introduced in the CPS in January 2003, at which point the harmonized estimate switches (discontinuously) to “Black alone”.<sup>22</sup>

22. The harmonized series is below the Census Bureau’s series because ours is an intercensal estimate that incorporates the 2010 Census, which was lower than estimated in the Vintage 2000 estimate.

We create population targets for 1990–2002 in three steps. The first step is to calculate the intercensal estimate for 2000–10 on a multi-race basis from the 2000 Census base, which is required because the postcensal estimates for 2000–02 do not incorporate revisions stemming from the 2010 Census. This ensures that the population from April 2000 to December 2002 is consistent with the 2010 Census estimates base.

The second step is to convert the intercensal estimate from step one to a single-race basis. The total for a race on a single race basis is equal to the number reporting that race alone, plus a fraction of those reporting that race in combination. (The next section describes how that fraction is calculated.) This provides the April 2000 population level that is consistent with the 1990 race categories.

Finally, the third step is to compute the error of closure for April 2000 on a single-race basis and create the intercensal estimate for 1990–2000.

#### A4.1. Estimating single-race categories for the 2000 Census base

From 2000 forward, the postcensal estimates by characteristics published by the Census Bureau report race for 5 race categories

- White
- Black or African American
- American Indian and Alaska Native
- Asian
- Native Hawaiian and Other Pacific Islander

The data include estimates for “ $R$  alone” and “ $R$  alone or in combination”, as well as a category for two or more races.

To make these categories historically comparable for the CPS population targets, we need to reclassify a fraction  $\theta_R$  of individuals in the “ $R$  in combination” category as in the “ $R$  alone” category.

The total for a race on a single-race basis is equal to the number reporting that race alone, plus a fraction of those reporting that race in combination. The assumption underlying our adjustment is that those reporting a combination would have been equally likely to report any of their races under the single race concept.

The number of people reporting a combination including race  $R$  is equal to the number reporting  $R$  either alone or in combination ( $rac$ ) minus the number reporting  $r$  alone ( $ra$ ). Summing this across races will double count these individuals, so the total should be roughly twice the number of people reporting two or more races. This ratio is  $\theta_R$ .

For each category  $R$ , the first-stage adjustment is

$$(A.2) \quad \delta_R = \max \left( 0, \frac{Rac}{Ra} - 1 \right),$$

which is the fraction of “ $R$  alone or in combination” that exceeds “ $R$  alone”. However, because individuals can select as many races as necessary, the first-stage adjustment may not sum to the “two or more” category. To account for this, define

$$(A.3) \quad \Theta = \frac{(wac - wa) + (bac - ba) + (iac - ia) + (aanac - aana)}{tom},$$

which accounts for the fact that the sum of the “in combination” races is greater than the total population.  $\Theta$  is a bit bigger than 2 (since every person who selects more than one race picks at least 2).

All told, the adjustment is

$$(A.4) \quad \theta_R = 1 + \frac{\delta_R}{\Theta}.$$

The ratios are calculated from the Vintage 2009 data (intercensal estimate, to account for errors from the 2010 Census) and is done separately for each month, sex, age, and ethnicity. After the adjustment, we sum the population to race and ethnicity categories that existed in the single race concept.

#### A5. 2000–2020 Census bases (January 2003 to December 2024)

Calculating the 2000–10 and 2010–20 intercensal estimates is straightforward, as the postcensal estimates for this period contain all the necessary information. For each census base, we calculate the error of closure for the final postcensal vintage of a census base and smooth that error back to April 1st of the previous decade using equation (A.1). Finally, for the current cen-

2020), we can simply use the latest postcensal estimate, currently Vintage 2024.

The table below lists the postcensal vintages used in our estimates:

Census base	Period covered	Postcensal data source
2000	January 2003–March 2010	<a href="#">Vintage 2009</a>
2010	April 2010–March 2020	<a href="#">Vintage 2020</a>
2020	April 2020–present	<a href="#">Vintage 2024</a>

## B. Calculating second stage weights

We take the published CPS second-stage weight as a starting point and reweight to match our harmonized population targets, following mostly the same procedure as the BLS uses to construct the original weights. Reweighting starts by matching national demographic totals at a disaggregated level (“national coverage step”), before iteratively matching multiple demographic targets along different dimensions (“second-stage weighting”). In our reweighting, we do not implement the state coverage step or the state/sex/age component of the second-stage weighting since the population targets for these steps are not publicly available.

### National coverage step

We adjust the second stage weights so that the totals by demographic groups match our harmonized population estimates. Records are grouped into four month-in-sample (MIS) pairs: MIS 1 and 5, MIS 2 and 6, MIS 3 and 7, and MIS 4 and 8. Each MIS pair is then adjusted to age/sex/race/ethnicity population controls using the following formula:

$$(B.1) \quad Y_{jkt}^{NC} = Y_{jkt}^{FS} \times \frac{C_{jt}}{E_{jkt}}$$

where

$Y_{jkt}^{FS}$  = sum of first-stage weights for month  $t$  in cell  $j$  and MIS pair  $k$ ,  
 $C_{jt}$  = national coverage adjustment control for month  $t$  in cell  $j$ ,  
 $E_{jkt}$  = weighted tally (using first-stage weights) for month  $t$  in cell  $j$  and MIS pair  $k$ .

The cells for the national coverage step are defined table B1. Some ages needed to be grouped together in order to maximize demographic detail while limiting cells with fewer than 20 persons responding each month. The degree of demographic detail varies over time based on the information available. Some noteworthy limitations that informed our choice of cells:

- We do not have independent population targets by Hispanic origin before April 1980
- Children’s records are not present in the public-use microdata files (PUMF) before January 1982

For 2003 forward, we use largely the same cells as the BLS reports in *Current Population Survey Design and Methodology* Technical Paper 77 (U.S. Census Bureau 2019, hereafter “CPS Technical Paper 77”). One difference is that we control ages 16 to 17 and ages 18 to 19 separately, whereas Table 2-3.1 has a single group for ages 16 to 19. The BLS did not target “Asian alone” as a separate race category until 2003, even though the CPS included “Asian or Pacific Islander” as a response option starting in 1994; we followed the BLS’s practice.

### Second-stage weighting

We next adjust the weights within each MIS pair such that the sample estimates for demographic subgroups are matched to our harmonized population controls. Two sets of controls are used:<sup>23</sup>

1. Ethnicity/sex/age: See table B2
2. Race/sex/age: See table B3

For 10 iterations, weights are updated to match each of these targets in turn. At each iteration, the new weights,  $w''_{i,t}$ , are calculated from the previous iteration’s weights  $w'_{i,t}$ :

$$(B.2) \quad w''_{i,t} = w'_{i,t} \left( \frac{p_{t,d}}{\sum_{i \in d} w'_{i,t}} \right) \quad \text{for } i \in d$$

---

23. We have not yet implemented the state/sex/age step.

**Table B1. National coverage step cell definitions**

Age/Period	White alone, non-Hispanic				White alone, Hispanic				Black alone, non-Hispanic				Asian alone, non-Hispanic				All other races, non-Hispanic				Non-White alone, Hispanic			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
0																								
1																								
2																								
3																								
4																								
5																								
6																								
7																								
8																								
9																								
10-11																								
12-13																								
14																								
15																								
16-17																								
18-19																								
20-24																								
25-29																								
30-34																								
35-39																								
40-44																								
45-49																								
50-54																								
55-59																								
60-62																								
63-64																								
65-69																								
70-74																								
75 and older																								

Note: Shaded cells indicate an independent population target. All targets are computed separately for male and female. Period 1 is January 1976 to March 1980; there is no distinction by ethnicity. Period 2 is April 1980 to December 1981. Period 3 is January 1982 to December 2002. Period 4 is January 2003 to the present.

Before 2003, the targeted race categories were “White” and “Black”; see section A4.

**Table B2. Second-stage adjustment cell definitions by ethnicity, age, and sex**

Age/Period	Hispanic				Non-Hispanic			
	1	2	3	4	1	2	3	4
0-1								
2-4								
5-7								
8-9								
10-11								
12-13								
14								
15								
16-19								
20-24								
25-29								
30-34								
35-39								
40-44								
45-49								
50-54								
55-64								
65 and older								

Note: Shaded cells indicate an independent population target. All targets are computed separately for male and female. Period 1 is January 1976 to March 1980; there is no distinction by ethnicity. Period 2 is April 1980 to December 1981. Period 3 is January 1982 to December 2002. Period 4 is January 2003 to the present.

using each of the definitions of demographic group cells  $d$ , with  $p_{t,d}$  representing our harmonized population estimate for group  $d$ . This iterative proportional fitting, or “raking”, procedure ensures that both sets of targets are matched as well as possible (Stephan 1942).

As with the national coverage step, the degree of demographic detail varies by time period. Some ages needed to be grouped together to avoid cells with fewer than 20 persons responding each month.

In particular, for 2003 forward, our cell definitions for the ethnicity/sex/age step are the same as in Table 2-3.3 of CPS Technical Paper 77. For the race/sex/age step we use the same cells for “White alone” but have aggregated a few cells in the “Black alone” and “All other races” categories relative to what is reported in Table 2-3.4 of CPS Technical Paper 77 because of insufficient observations in the PUMF.

**Table B3. Second-stage adjustment cell definitions by race, age, and sex**

Age/Period	White alone				Black alone				All other races			
	1	2	3	4	1	2	3	4	1	2	3	4
0												
1												
2												
3												
4												
5												
6												
7												
8												
9												
10–11												
12–13												
14												
15												
16–17												
18–19												
20–24												
25–29												
30–34												
35–39												
40–44												
45–49												
50–54												
55–59												
60–62												
63–64												
65–69												
70–74												
75 and older												

Note: Shaded cells indicate an independent population target. All targets are computed separately for male and female. Period 1 is January 1976 to March 1980. Period 2 is April 1980 to December 1981. Period 3 is January 1982 to December 2002. Period 4 is January 2003 to the present.

Before 2003, the targeted race categories were “White” and “Black”; see section A4.

## C. Calculating composite weights

The composite estimators for employment and unemployment take an *AK* form (Breau and L. Ernst 1983), combining estimates in levels and changes:

$$(C.1) \quad \hat{Y}_t^{\text{Composite}} = (1 - K) \underbrace{\hat{Y}_t^{\text{Second stage}}}_{\text{Level estimate}} + K \underbrace{\left( \hat{Y}_{t-1}^{\text{Composite}} + \Delta_t \right)}_{\text{Previous + change est.}} + A \underbrace{\hat{\beta}_t}_{\text{Incoming group chg.}}$$

The standard estimate in levels is given by:

$$(C.2) \quad \hat{Y}_t^{\text{Second stage}} = \sum w_{i,t}^{\text{Second stage}} 1(\text{Labor force status}(i) = Y)$$

An alternative estimate is based on the change among continuing rotation groups:

$$(C.3) \quad \Delta_t = \frac{4}{3} \left( \sum_{\substack{\text{MIS}(i) \in \\ \{2-4, 6-8\}}} w_{i,t}^{\text{SS}} 1(\text{LFS}(i) = Y) - \sum_{\substack{\text{MIS}(i) \in \\ \{1-3, 5-7\}}} w_{i,t-1}^{\text{SS}} 1(\text{LFS}(i) = Y) \right)$$

An additional adjustment is included for the difference between incoming and incumbent rotation groups:

$$(C.4) \quad \hat{\beta}_t = \sum_{\text{MIS}(i) \in [1,5]} w_{i,t}^{\text{SS}} 1(\text{LFS}(i) = Y) - \frac{1}{3} \left( \sum_{\substack{\text{MIS}(i) \in \\ [2-4, 6-8]}} w_{i,t}^{\text{SS}} 1(\text{LFS}(i) = Y) \right)$$

We follow the BLS in using  $A = 0.3$  and  $K = 0.4$  for the level of unemployment, and  $A = 0.4$  and  $K = 0.7$  for the level of employment (CPS Technical Paper 77).

The composite weighting process proceeds iteratively, similar to the raking process to compute the second-stage weights. Within each labor category, a three-dimensional rake is applied, using the second-stage estimate at the state level as one step and national second-stage estimates for age/sex/race (table C1) and age/sex/ethnicity (table C2) as the second and third steps.<sup>24</sup>

24. For state-level estimates, California and New York are split into two parts and each part is treated like a state. California is split into Los Angeles County and the rest of Cal-

**Table C1. Composite national race cell definitions**

Age/Period	White alone				Black alone				All other races			
	1	2	3	4	1	2	3	4	1	2	3	4
16–19												
20–24												
25–29												
30–34												
35–39												
40–44												
45–49												
50–54												
55–59												
60–64												
65 and older												

Note: Shaded cells indicate a composite target for employment and unemployment. All targets are computed separately for male and female. Period 1 is January 1976 to December 1988. Period 2 is January 1989 to December 2002. Period 3 is January 2003 to December 2018. Period 4 is January 2019 to the present.

Before 2003, the targeted race categories were “White” and “Black”; see section A4.

**Table C2. Composite national ethnicity cell definitions**

Age	Hispanic	Non-Hispanic
16–19		
20–24		
25–34		
35–44		
45 and older		

Note: Shaded cells indicate a composite target for employment and unemployment. All targets are computed separately for male and female.

## D. Time series

### D1. Adjustments for the CPS redesign

As Polivka and Miller (1998, page 249) detail, in January 1994, the CPS underwent a major redesign both in the wording of the questionnaire and the methodology used to collect the data. The objective of the redesign was to improve the quality and expand the quantity of available data. However, the redesign also caused changes in the measurement of many of the estimates derived from the CPS. Polivka and Miller estimate adjustment factors

ifornia. New York is split into New York City (New York, Queens, Bronx, Kings, and Richmond Counties) and the rest of New York. The District of Columbia is treated as a state.

for various aggregate measures derived from the CPS in order to permit comparisons of estimates before and after the redesign.

We apply their adjustment factors to our harmonized time-series from 1948–93, so that the resulting series are comparable to the post-redesign measures. We calculate labor force statistics for 30 groups (15 age groups  $\times$  2 genders). Finally, we use the multiplicative factors for the labor force participation rate (LFPR) and the employment-to-population ratio (EPOP) ratio, by demographic group, to adjust the levels of the labor force and employment. The other labor force stocks are calculated by subtraction.

## D2. Extending the time series back to 1948

Since our methodology for estimating harmonized labor force statistics uses the CPS microdata, the paper’s main estimates start in January 1976, the first month for which there are PUMF. However, there are also breaks in the time series before 1976, arising from when the BLS incorporated updated population estimates, but they occur less frequently than today because the external population controls were typically updated only once per decade. However, the breaks that are present in the pre-1976 series tend to be sizable and affect the composition of the population across demographic groups.<sup>25</sup>

The key breaks in the time series before 1976 are listed in table D1. The BLS reported the effect of the introduction of the updated population controls in its monthly labor report. We collected the data from those reports and used those to smooth out the breaks.

**Table D1. Noncomparability of population and labor force levels**

Date of break	Reason	Demographic detail
January 1953	Introduction of 1950 Census	Sex/age
January 1960	Addition of Alaska and Hawaii	n.a.
April 1962	Introduction of 1960 Census	Sex/age/race
January 1972	Introduction of 1970 Census	Sex/age/race
January 1974	Revised postcensal methodology	Sex/age/race

25. For additional information, see “Appendix: History of the Current Population Survey” beginning on page 29 of U.S. Census Bureau (2019) and the section “Noncomparability of labor force levels” beginning on page 197 of the February 2006 issue of *Employment and Earnings* (Bureau of Labor Statistics 2006).

Our harmonized series adjust for all breaks except January 1960, when Alaska and Hawaii were added to the CPS sample. Although their addition creates a one-time shift up in the population, we did not smooth this break backwards, as it reflects a legal change in the United States population rather than a statistical artifact. (We recognize, however, that this distinction is not entirely bright.) We start by smoothing back the January 1953 revision and apply each subsequent adjustment in chronological order to the previously adjusted time series.

The starting point for our harmonized time series are the published series for the CNP, the civilian labor force, and civilian employment by sex (male and female) and by 8 age groups (16 to 17, 18 to 19, 20 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 and older). Time series for the White race category begin in January 1954 and those for the Black race category begin in January 1972. Unfortunately, there is no information on the effects of the updated population controls by race, so we use each race's share of the relevant level to allocate the revision.

**January 1953** Updated population estimates drawn from 1950 Census were introduced in January 1953. Population levels were raised by about 600,000, while the labor force and employment were increased by about 350,000, with the primarily affecting the figures for totals and for men; other categories were relatively unaffected. The effects of the updated population controls for population and the labor force in 16 sex/age cells are reported in *Current Population Reports*. Unfortunately, the article does not include information about the effect on employment, so we adjusted employment such that the EPOP ratio in December 1952 was unaffected. The resulting revisions are smoothed back linearly to April 1940. In practice, however, because the published data begin in January 1948, the adjustment is not neutral for the starting level (see figure D1).

**April 1962** Updated population estimates drawn from 1960 Census were introduced in April 1962. The population level was reduced about 50,000, while the labor force was reduced about 200,000 and employment a bit less; unemployment totals were virtually unchanged. The effects of the updated population controls for population and the labor force in 16 sex/age cells are reported in the May 1962 issue of *Employment and Earnings* (Bu-

reau of Labor Statistics 1962). The effects on employment are reported by sex and age group, but the youngest age group is 14 to 19. We allocate the employment revision for this group into 14 to 15, 16 to 17, 18 to 19 based on each group's share of employed using the updated population controls. The resulting revisions are smoothed back linearly to April 1950.

**January 1972** Updated population estimates drawn from 1970 Census were introduced in January 1972. The effects of the updated population controls for the population, labor force, and employment in 14 sex/age cells are reported in the February 1972 issue of *Employment and Earnings* (Bureau of Labor Statistics 1972). The youngest age group reported is 16 to 19, so we allocate the revisions for this group into 16 to 17 and 18 to 19 based on each group's share using the updated population controls. The resulting revisions are smoothed back linearly to April 1960.

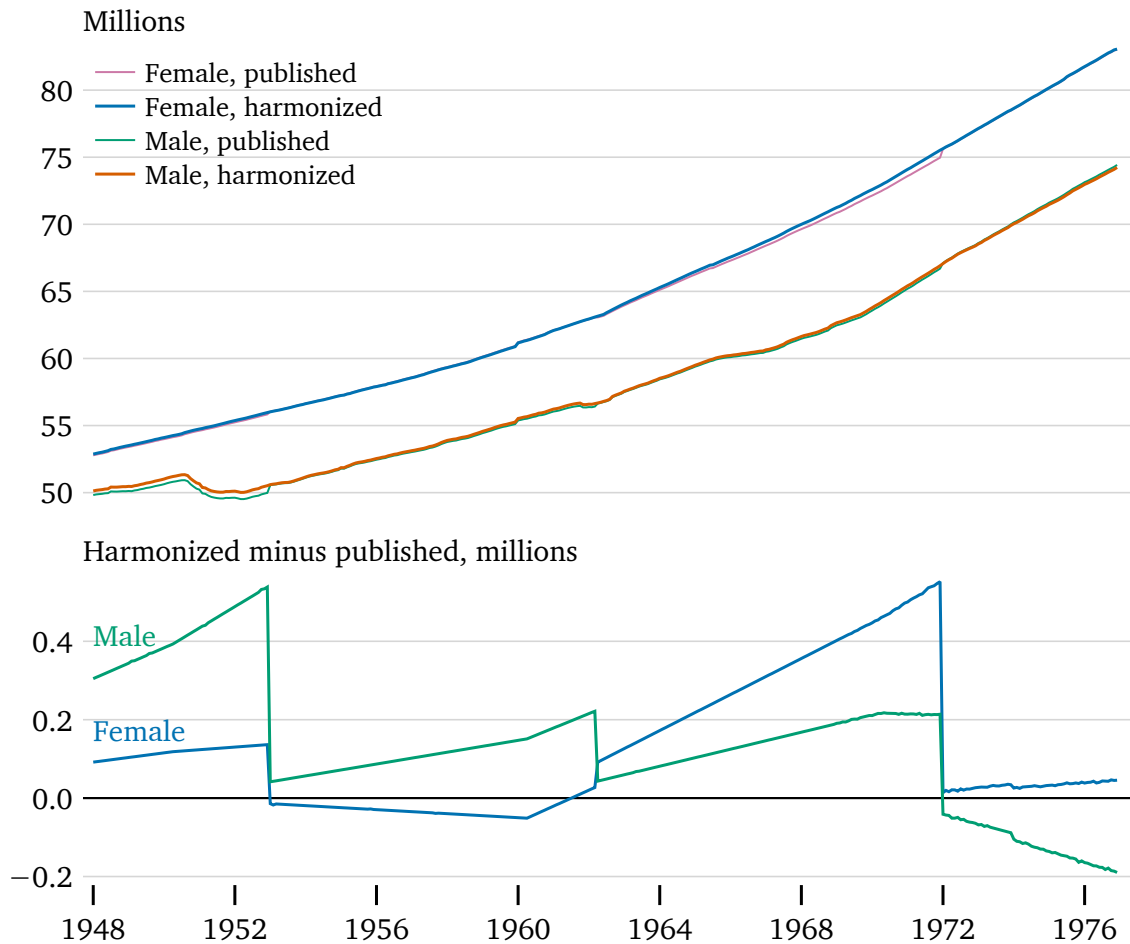
**January 1974** Beginning in January 1974, the Census Bureau how it estimated the postcensal population to better account for the 1970 Census undercount.<sup>26</sup> The methodology change had the greatest impact on estimates of men aged 20 to 24, particularly in the non-White population, but had little effect on estimates of the total population aged 16 or older. The effects of the updated population controls for the population, labor force, and employment are reported for 14 sex/age cells. The youngest age group reported is 16 to 19, so we allocate the revisions for this group into 16 to 17 and 18 to 19 based on each group's share using the updated population controls. The resulting revisions are smoothed back linearly to April 1970, as this methodology change affected only the postcensal estimates.

**January 1976** Beginning in January 1976, our harmonized time series are calculated from the CPS PUMF. For all sex/age/race groups, we smooth any difference in January 1976 level between the microdata estimate and the adjusted historical time series back to April 1970.

---

26. See "CPS Population Controls Derived from Inflation-Deflation Method of Estimation" on page 7 of the February 1974 issue of *Employment and Earnings* (Bureau of Labor Statistics 1974).

**Figure D1. Civilian noninstitutionalized population, harmonized and published, 1948–76**



Note: Individuals aged 16 or older.

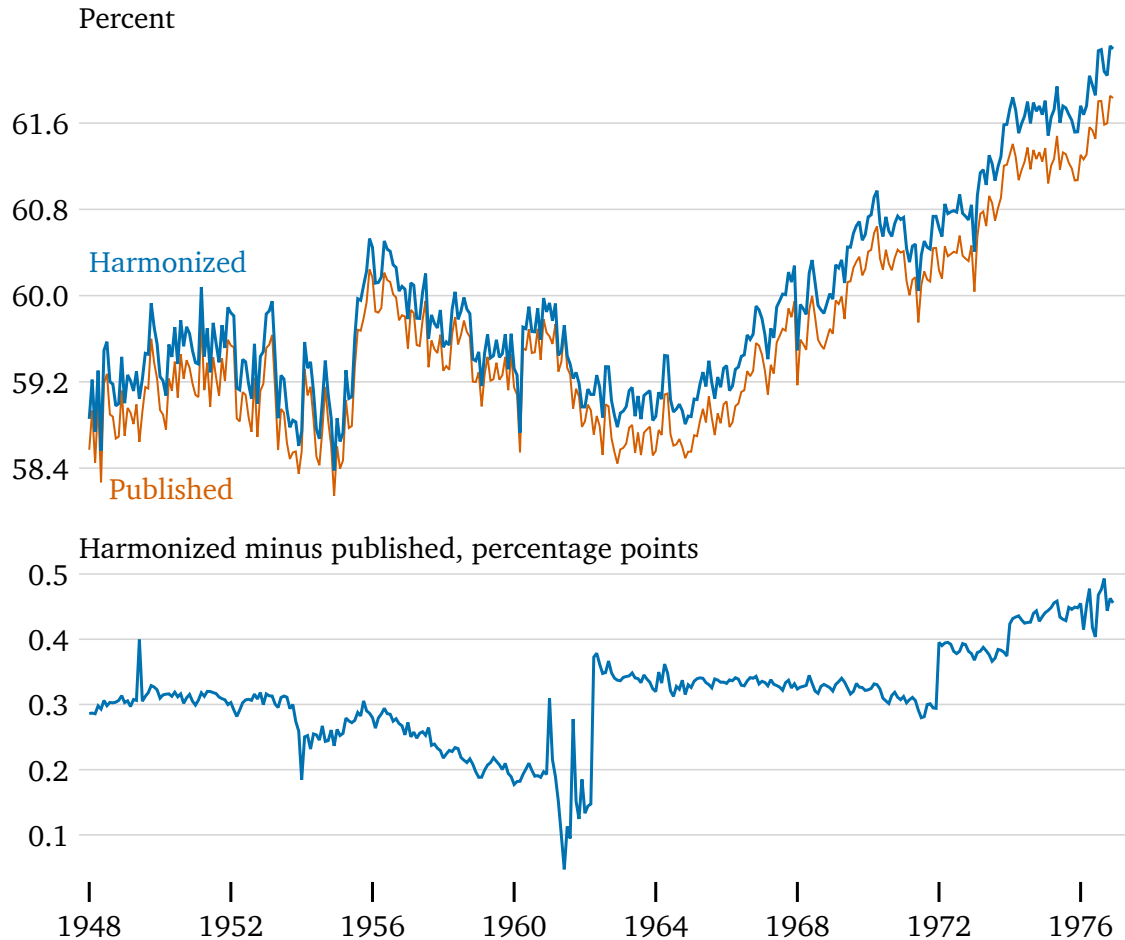
Source: Bureau of Labor Statistics (BLS); authors' calculations using data from the Census Bureau and the BLS.

### D2.1. Harmonized time series, 1948–76

The top panel of figure D1 plots the harmonized population for women and men over 1948–76, while the lower panel shows the difference between two series. The data for 1948–75 are the adjusted published time series. From January 1976 forward, the data are calculated from the CPS PUMF.

As shown by the green line in the lower panel, the harmonized male population is consistently above the published series for most of the period, often exceeding 200,000 and peaking at over 400,000 in December 1952. The difference for men turns negative starting in 1972 and falls almost

**Figure D2. Labor force participation rate, harmonized and published, 1948–76**



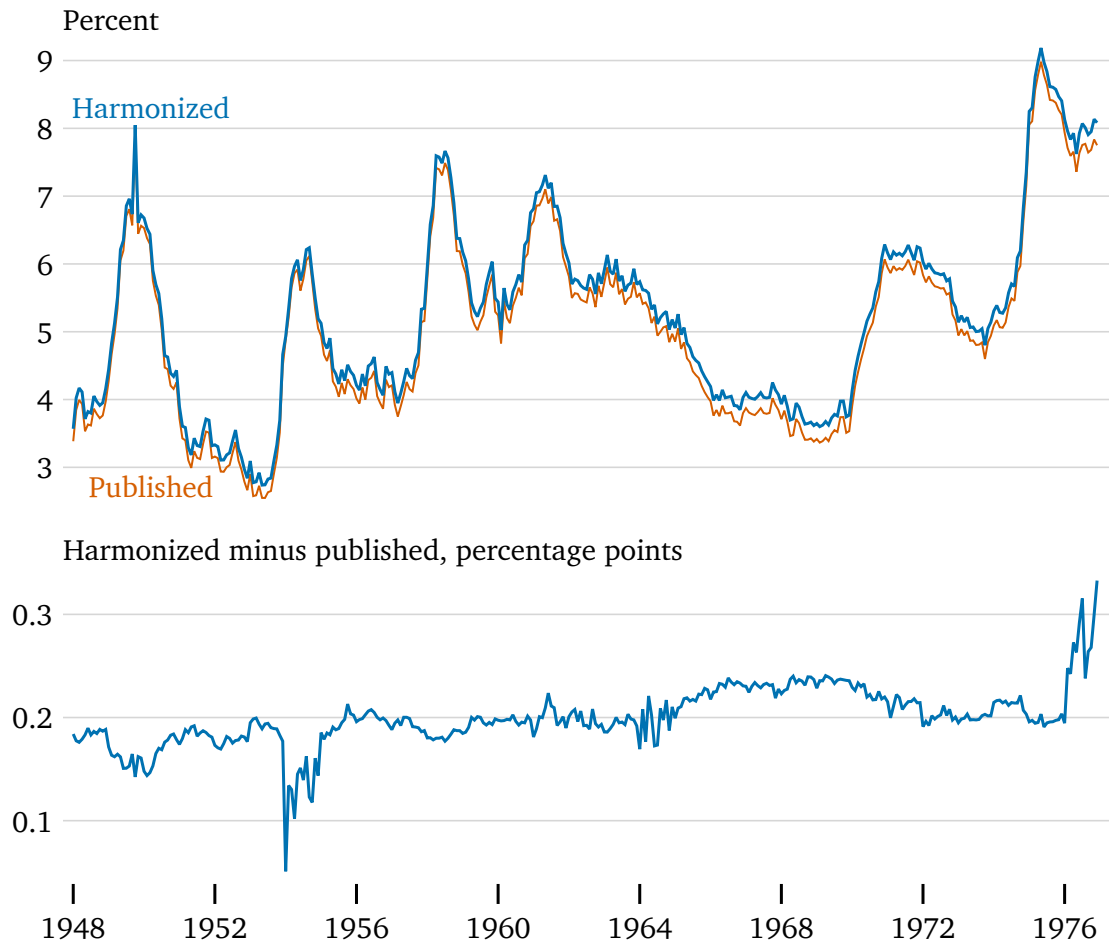
Note: Individuals aged 16 or older.

Source: Bureau of Labor Statistics (BLS); authors' calculations using data from the Census Bureau and the BLS.

–200,000 by January 1976. As shown by the blue line, the harmonized female population is slightly higher than the published series before 1953. The difference drops close to or slightly below zero through April 1962 before climbing to nearly 500,000 in December 1971.

Figures D2 to D4 show the effects of our harmonized data on the LFPR, unemployment rate, and the EPOP ratio. The blue line in the lower panel of figure D2 shows that the harmonized LFPR generally runs about 0.3 percentage point (pp) above the published LFPR series over 1948–76. The majority of this differential reflects the adjustment for the CPS redesign in 1994 (see figure 5 and the discussion in section 5.1).

**Figure D3. Unemployment rate, harmonized and published, 1948–76**



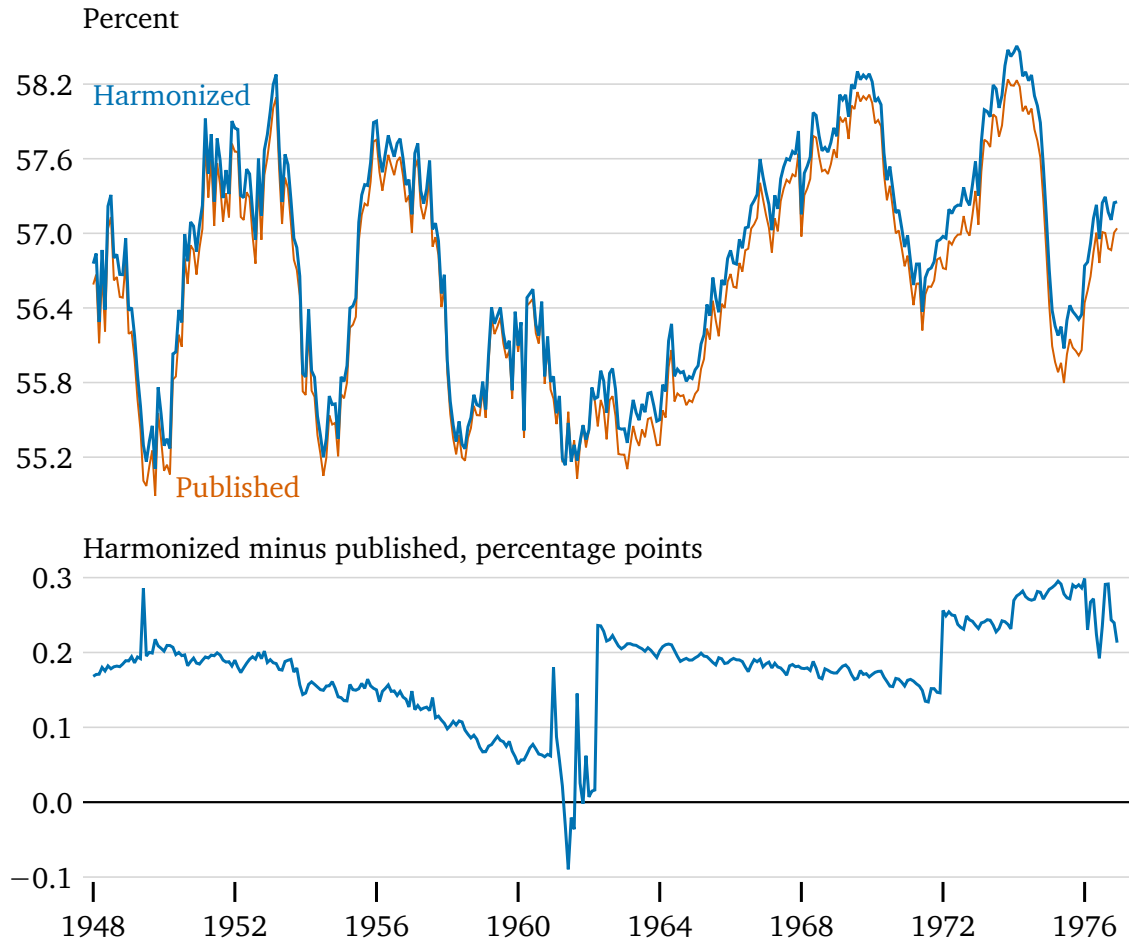
Note: Individuals aged 16 or older.

Source: Bureau of Labor Statistics (BLS); authors' calculations using data from the Census Bureau and the BLS.

The harmonized unemployment rate (figure D3) runs about 0.2 pp above the published series and this difference is relatively stable over time. The level shift is essentially “inherited” from the CPS period (see figure 6 and the discussion in section 5.1).

The positive differences between the harmonized and published series in both the LFPR and the unemployment rate partially offset for the EPOP ratio (figure D4).

**Figure D4. Employment-to-population ratio, harmonized and published, 1948–76**



Note: Individuals aged 16 or older.

Source: Bureau of Labor Statistics (BLS) and authors' calculations using data from the Census Bureau and the BLS.

### D3. Validating our replication

To verify that our methodology reproduces the BLS's procedures, we recreate the BLS' special tabulations of the December 2024 data that incorporate the new population controls. Since our methodology uses the latest population estimates, our harmonized series should match these special tabulations for December.

Table D2 reports the population for selected demographic groups, as well as the levels the labor force and employment. The revisions to population for nearly all groups match to within rounding. (We calculate totals as the sum of separate estimates for female and male, thus a difference of  $\pm 2$  is at the

precision of rounding.) Our population level for Asians is 0.2 percent lower than in the BLS's tabulation; this difference likely arises because the Asian race category is not targeted independently in the second-stage coverage steps.

For the labor force and employment levels our estimates are quite close to the BLS's tabulations, generally less than 0.1 percent different (in absolute value) — and less than 0.05 percent for the total. These small differences may be due to the effects of privacy protections in the CPS microdata, particularly perturbing ages in the published microdata. (The published revision table is calculated based on unperturbed microdata used internally by BLS, while our estimates use the perturbed public data.)

The differences also likely reflect differences in the demographic groups that we use when calculating our second-stage and composite weights. Indeed, the largest discrepancy in percentage terms is for both sexes aged 16 to 19, where our labor force and employment estimates are more than 0.6 percent below the BLS's. We control to ages 16 to 17 and ages 18 to 19 separately in the national coverage step, whereas the BLS has a single group for ages 16 to 19, which allows us to more closely match the population targets within the 16 to 19 age group. Because participation rates tend to rise sharply with age, this may contribute to the relatively large — though still small in absolute terms — difference between our estimate and the BLS's.

#### **D4. Additional results**

The top panel of figure D5 shows the LFPR for men and women and the bottom panel shows the difference between the harmonized estimate and the published series. The differences between harmonized and published before 1994 are larger by sex than for the topline participation rate, owing to a larger adjustment to women's participation rates from the CPS redesign than for men's.

Figure D6 shows the joint implications of the LFPR (figure 5) and unemployment rate (figure 6) for the EPOP ratio. The difference between the harmonized and published EPOP ratios in 1976 was smaller than that for the LFPR alone, because the harmonized unemployment rate was also higher than published. Thus, although the labor force was higher than published,

**Table D2. Population and labor force status for selected groups, December 2024**

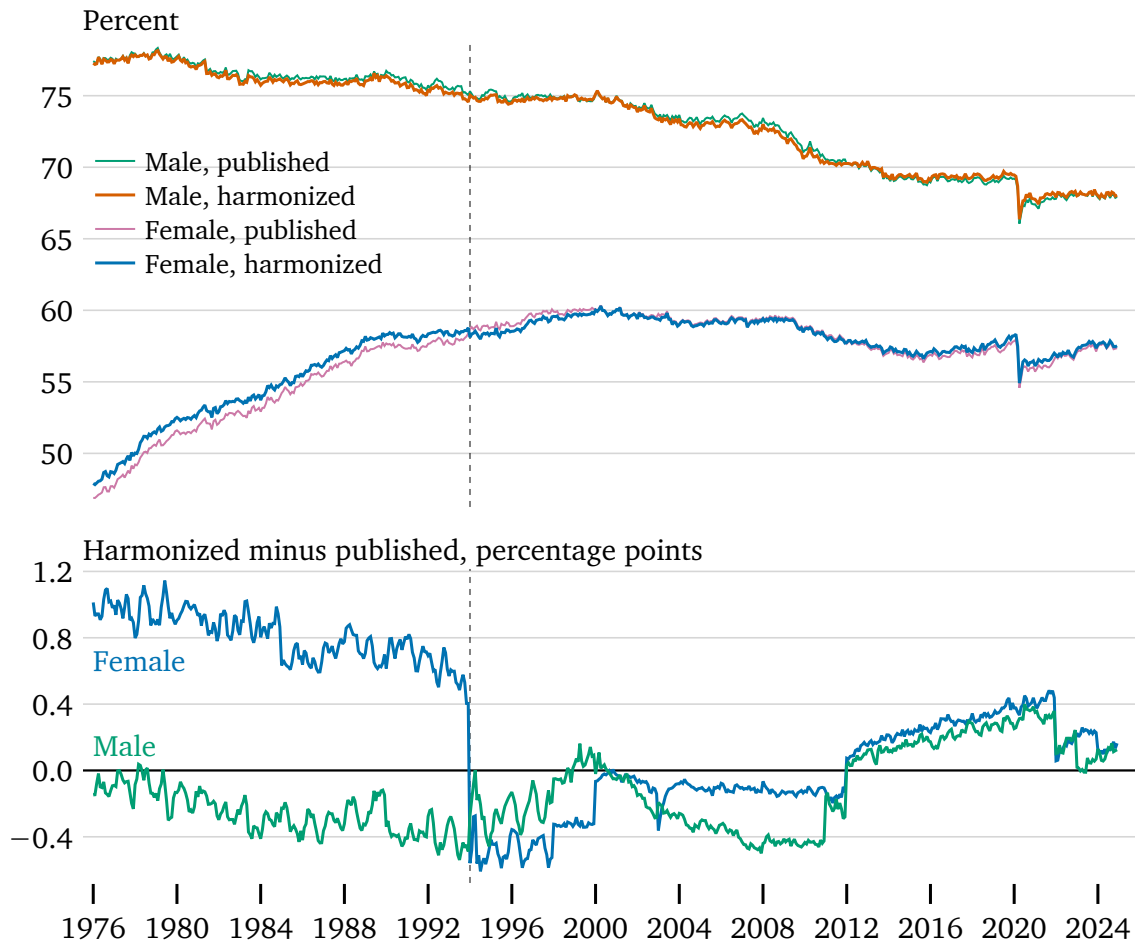
Thousands

Employment status	Harmonized	Updated	Difference
<b>TOTAL</b>			
Civilian noninstitutional population	272,509	272,509	0
Civilian labor force	169,932	169,852	80
Employed	163,366	163,294	72
<b>Men, 16 years and over</b>			
Civilian noninstitutional population	132,925	132,925	0
Civilian labor force	89,869	89,868	1
Employed	86,240	86,235	5
<b>Men, 20 years and over</b>			
Civilian noninstitutional population	123,834	123,833	1
Civilian labor force	86,708	86,676	32
Employed	83,458	83,442	16
<b>Women, 16 years and over</b>			
Civilian noninstitutional population	139,583	139,583	0
Civilian labor force	80,063	79,983	80
Employed	77,126	77,059	67
<b>Women, 20 years and over</b>			
Civilian noninstitutional population	130,819	130,817	2
Civilian labor force	76,971	76,879	92
Employed	74,327	74,235	92
<b>Both sexes, 16 to 19 years</b>			
Civilian noninstitutional population	17,856	17,858	-2
Civilian labor force	6,253	6,297	-44
Employed	5,581	5,616	-35
<b>WHITE</b>			
Civilian noninstitutional population	207,017	207,017	0
Civilian labor force	128,537	128,440	97
Employed	124,107	124,026	81
<b>BLACK OR AFRICAN AMERICAN</b>			
Civilian noninstitutional population	35,586	35,586	0
Civilian labor force	22,086	22,098	-12
Employed	20,844	20,850	-6
<b>ASIAN</b>			
Civilian noninstitutional population	18,972	19,018	-46
Civilian labor force	12,177	12,207	-30
Employed	11,770	11,800	-30
<b>HISPANIC OR LATINO ETHNICITY</b>			
Civilian noninstitutional population	50,761	50,760	1
Civilian labor force	34,238	34,243	-5
Employed	32,497	32,504	-7

Note: Not seasonally adjusted. Estimates for the above race groups (White, Black or African American, and Asian) do not sum to totals because data are not presented for all races. Persons whose ethnicity is identified as Hispanic or Latino may be of any race.

Source: Bureau of Labor Statistics (BLS) and authors' calculations using data from the Census Bureau and the BLS.

**Figure D5. Labor force participation rate, by sex, harmonized and published**



Note: Individuals aged 16 or older. Vertical line denotes January 1994, when the CPS redesign was implemented.

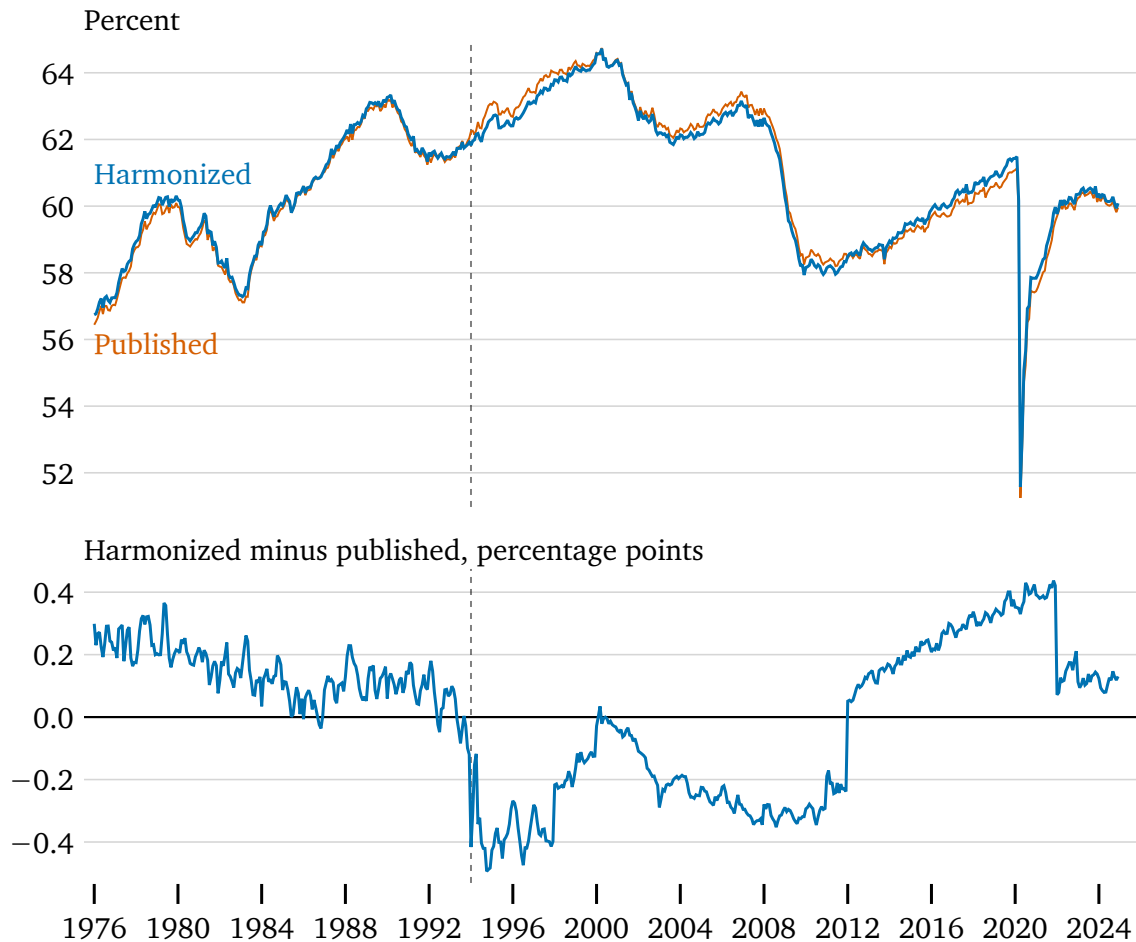
Source: Bureau of Labor Statistics (BLS) and authors' calculations using data from the Census Bureau and the BLS.

somewhat more of that higher labor force was also unemployed, mitigating the boost to employment. The difference between harmonized and published diminishes to about a tenth, on average, from the mid-1980s through the early 1990s.

## E. Comparison with BLS experimental series on smoothed labor force statistics

As we have discussed, the official time series for population and labor force statistics have a large discontinuity in January 2025, when the BLS began

**Figure D6. Employment-to-population ratio, harmonized and published**



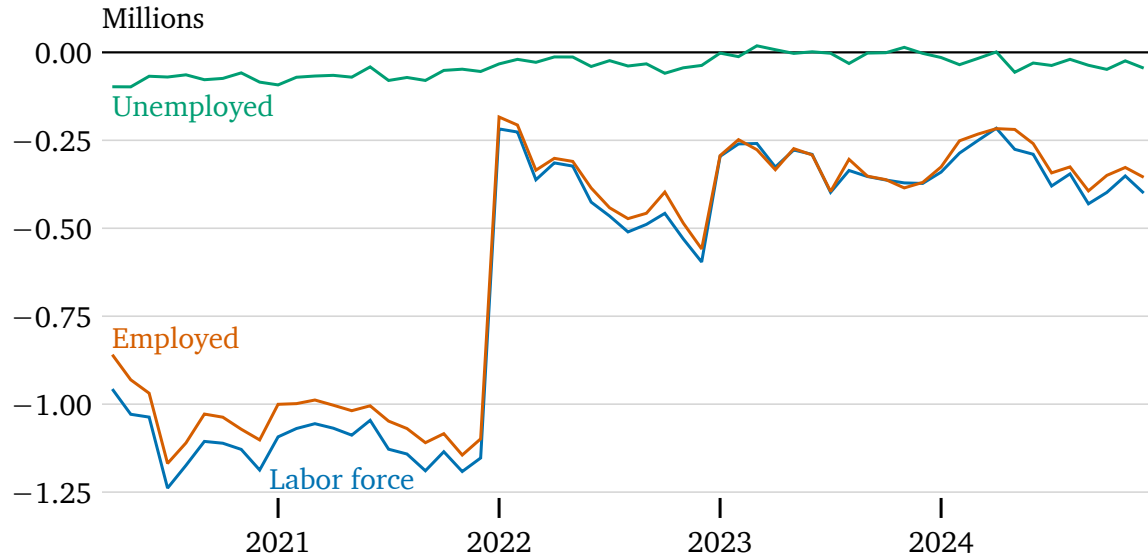
Note: Individuals aged 16 or older. Vertical line denotes January 1994, when the CPS redesign was implemented.

Source: Bureau of Labor Statistics (BLS) and authors' calculations using data from the Census Bureau and the BLS.

using the Census Bureau's V2024 population estimates. The BLS also released a set of experimental time series that smooth out this break.

These experimental series scale up the published time series for the labor force, employment, and unemployment from April 2020 to December 2024 by the ratio of that month's V2024 estimate of the CNP population (aged 16 or older) to the BLS's official population series, which reflects the population estimate as of the time of original publication. However, such a proportional adjustment means that these revisions do not affect the demographic

**Figure E1. Comparison with BLS experimental labor force statistics**



Note: Bureau of Labor Statistics (BLS) experimental series minus harmonized estimate. Not seasonally adjusted.

Source: Authors' calculations using data from the BLS.

composition of the population.<sup>27</sup> In other words, this adjustment assumes that the demographic composition of each month's revisions to the population are the same as in the older vintages of the population estimates that were used to produce the original labor force statistics for each month.

As our methodology harmonizes the labor force statistics by targeting the latest population estimates for the detailed demographic groups used by the BLS to generate its CPS weights, our harmonized labor force statistics account for both revisions to the aggregate CNP population and revisions to the demographic composition of the population. As a result, there are non-trivial differences between our harmonized labor force statistics and the BLS's experimental labor force statistics.

Figure E1 shows the difference between the BLS's experimental estimate and our harmonized estimate for the labor force (the blue line), employed (the orange line) and unemployed (the green line). The sizable gaps between these estimates from April 2020 through December 2021 shrink considerably in January 2022 when the published BLS statistics switch to the

27. The BLS notes this feature in its documentation of the experimental series, in which it also pointed to considerable uncertainty in the demographic composition of the humanitarian migrants (Bureau of Labor Statistics 2025).

Vintage 2022 (V2022) population estimates. The V2022 estimates were the first to reflect the 2020 Census’s revisions to the demographic composition as of April 2020 — which revised down the population of older individuals (who tend to have lower participation and employment rates) and revised up the prime-age population (who tend to have higher participation and employment rates). From January 2022 forward, both the published BLS labor force statistics and our harmonized labor force statistics reflect the revisions from the 2020 Census, thus considerably shrinking the gap between our estimates. However, some gap between our estimates remains through December 2024 because our harmonized labor force statistics reflect the evolution of the demographic composition of the population implied by the V2024 population estimates, whereas the BLS’s experimental labor force statistics reflect the V2022 and Vintage 2023 (V2023) demographic compositions of the population for its labor force statistics in 2022 and 2023.

## **F. Incorporating administrative data on immigration**

This section relies heavily on information from the Census Bureau’s methodology for constructing population estimates (U.S. Census Bureau 2024b).

Net immigration is one of the three components of change that the Census Bureau separately estimates when constructing its estimate of population growth (with births and deaths being the other two components). Foreign-born immigration — that is, moves by noncitizens into the United States from a foreign country in which a change of usual residence has occurred — is one component of net immigration. Annual immigration totals are estimated separately for persons migrating from Mexico and for persons migrating from all other countries.

The Census Bureau then uses “proxy universes” to distribute the totals to single years of age, sex, race, and Hispanic origin. A proxy universe is a subset of the total ACS population that is used to represent the geographic and demographic composition of international migrants. Proxy universes provide considerably larger sample sizes than the input data used to estimate immigration totals. Annual immigration totals are distributed to national characteristics using a proxy universe pooled from the corresponding ACS 1-year file and two previous ACS 1-year files. The proxy universe for immi-

gration from Mexico is the ACS population born in Mexico whose year of entry into the United States was five years ago or less. The proxy universe for immigration from all other countries is the ACS population born in a foreign country other than Mexico whose year of entry was five years ago or less.

Since the start of the COVID-19 pandemic there have been considerable shifts in the dynamics of foreign-born immigration. Due to limitations in the timeliness and coverage of the ACS data that the Census Bureau relies on to estimate foreign-born immigration, these large shifts in immigration have often not been immediately apparent in the Census Bureau’s net immigration estimates (nor, as a result, in its population growth estimates).<sup>28</sup> Instead, in order to capture these shifts, the Census Bureau has had to change its methodology for estimating net immigration in subsequent vintages — augmenting the ACS data with administrative data on “humanitarian migration”.

These methodological adjustments have resulted in sizable revisions to the estimates of net immigration in prior years, which translate into significant revisions to the population. (Note that for the PEP’s population estimates, a “year” refers to an estimate year, which runs from July 1 of a calendar year to June 30 of the next calendar year.) For instance, the Census Bureau’s V2024 population estimates revised up the contribution of net immigration to annual population growth by nearly 700,000 in estimate year 2022 and by 1.2 million in 2023 — with nearly all of these revisions due to humanitarian migration. As prior years’ CPS labor force statistics rely on CPS micro-data weights that reflect only the outdated population estimates from previous years, these revisions to net immigration indicate possible distortions in the published labor force statistics from previous years.

Our baseline methodology described in section 4 eliminates these distortions in the labor force statistics by constructing harmonized CPS micro-data weights that reflect the accumulated revisions to the population es-

---

28. The Census Bureau noted these limitations in its documentation of the changes to the methodology used to construct the V2024 net immigration estimates (Gross et al. 2024). First, the Census Bureau noted that each year, when it produces a new vintage of its population estimates (and the components of change), it must rely on ACS data that was collected for the year prior. And second, the Census Bureau noted that some foreign-born populations, particularly recently arriving immigrants, are not well represented in the ACS.

timates for any given year. For instance, now that the Census Bureau has adjusted its V2024 methodology to better capture these shifts in net immigration, our harmonized CPS microdata weights can be used to remove the distortions in the labor force statistics resulting from the failure of earlier vintages of the population estimates to properly capture these shifts in net immigration.

In some cases, however, revisions may be predictable before the Census Bureau publishes a new vintage of its population estimates. For instance, almost a full year before the V2024 population estimates significantly revised up the pace of net immigration in 2022 and 2023, the Congressional Budget Office (2024) published population estimates with similarly large revisions to net immigration using many of the same administrative data sources that the V2024 methodology would ultimately rely on.

When population revisions, and their demographic composition, can be estimated, our methodology can be extended to generate CPS microdata weights and harmonized labor force statistics that reflect that alternative population. We use administrative data on migrant inflows to demonstrate how our methodology can harmonize CPS microdata weights and labor force statistics to reflect the Census Bureau's anticipated revisions to estimated net immigration in advance of their publication.

#### **F1. Census Bureau's adjusted methodology for estimating foreign-born immigration**

Both prior to the pandemic and in its V2023 population estimates, the Census Bureau estimated foreign-born immigrant inflows solely from ACS.<sup>29</sup> The V2023 estimates indicated that net immigration boosted population growth by nearly 1.2 million in 2023, a boost similar to net immigration's contribution to population growth prior to the pandemic.

However, population estimates published at about the same time by the Congressional Budget Office (CBO) indicated a much faster pace of net immigration in recent years, with the CBO estimating that net immigration

---

29. As with the Vintage 2021 and V2022 population estimates, the V2023 population estimates continued to use administrative data to adjust net immigration estimates for 2020 and 2021 to reflect pandemic-related travel restrictions. However, the V2023 estimates of net immigration for 2022–24 returned to relying primarily on data from the 2022 ACS.

reached 3.3 million in calendar year 2023 (Congressional Budget Office 2024) — an upward revision of almost 2 million to the CBO’s projected pace of 2023 net immigration from just a year prior. This upward revision resulted from the CBO using administrative data sources, which showed migrant inflows rising significantly starting in mid-2021.

For the V2024 population estimates, the PEP adjusted its methodology for estimating foreign-born immigration, incorporating data on migrant inflows from many of the same sources as the CBO. Humanitarian migration is calculated as 75 percent of immigrant inflows from the following categories:<sup>30, 31</sup>

1. Migrants encountered by U.S. Border Patrol (USBP) and released into the United States with a notice to appear before an immigration judge or granted parole at a port-of-entry by the Customs and Border Protection (CBP) Office of Field Operations (OFO),<sup>32</sup> and
2. Refugees admitted into the country.<sup>33</sup>

For estimate years 2021–24, annual counts of humanitarian migration are used to inflate the ACS-based foreign-born immigration and then the resulting totals are distributed to characteristics using the proxy universes.

## F2. Estimating alternative population targets

Nearly all the revision to the Census Bureau’s population estimates from V2023 to V2024 can be attributed to revisions in estimated net immigration. Thus, we can achieve a close approximation of the V2024 estimates if we assume that the V2023 population estimates are accurate for the native-

---

30. The PEP include only 75 percent of the totals from the administrative data because it believes that about  $\frac{1}{4}$  of the inflows are already captured in the ACS migration totals.

31. The CBO’s net immigration estimates also incorporated data on the number of unaccompanied minors who were apprehended when crossing the border and transferred to the custody of the Department of Health and Human Services, individuals who overstayed temporary visas, and migrants who entered the United States without encountering a Customs and Border Protection (CBP) official. The Census Bureau’s V2024 estimates do not incorporate these sources.

32. Monthly counts of migrants released by USBP or granted parole by the CBP OFO obtained from Department of Homeland Security, Office of Homeland Security Statistics (2025).

33. Monthly counts of refugees arriving in the United States by country of origin obtained from Department of State, Refugee Processing Center (2025).

**Table F1. Geography scheme for humanitarian migration**

Thousands

Name	Encounters per year
Central America	1,407
South America	587
Caribbean	272
Asia	139
Europe	57
Africa	56
North America	25
Oceania	1

Note: Average encounters per year for 2022–24.

Source: Department of Homeland Security, Office of Homeland Security Statistics (2025); Department of State, Refugee Processing Center (2025).

born population and focus on estimating the revision to net immigration from V2023 to V2024.

Like U.S. Census Bureau (2024b), we first calculate the revisions to foreign-born immigration and then distribute the totals to demographic characteristics using proxy universes. However, we depart from the V2024 methodology in the geographic detail used. The Census Bureau calculates immigration totals and distributes those totals to characteristics using two geographies: Mexico and all other countries. Because the 2022–24 immigration surge was heavily concentrated among Latin American countries, we use a more detailed geographic scheme for estimating foreign-born immigration and constructing proxy universes.

In particular, both the ACS and the administrative data on migrant flows have information about migrants' country of origin or citizenship. In order to ensure sufficient sample sizes, we group countries together into 8 regions based on the [United Nations M49 definitions](#) (Department of Economic and Social Affairs, Statistics Division 1999). We grouped countries in the Americas by intermediate region: Caribbean, Central America, South America, and North America. Countries outside the Americas are grouped by region: Africa, Asia, Europe, Oceania. Table F1 reports our geographic scheme, sorted by the average number of encounters per year over 2022–24. Encounters with citizens of Mexico, which is included in the Central America region, averaged 746,000 per year, easily the single biggest source country.

Our estimate of the revision to foreign-born immigration from V2023 to V2024 incorporates two sources of revision: (1) foreign-born immigration totals estimated from the ACS and (2) adding humanitarian migration.

On the first source, there are two revisions to the ACS-based foreign-born immigration totals:

1. **2021:** In the V2023 estimates, foreign-born immigration for 2021 was calculated as 103 percent of the 2019 ACS total. In the V2024 estimates, the PEP used the foreign-born immigration totals from the 2022 ACS. The revision for 2021 is the difference between these two values.
2. **2022–2024:** Due to the timing of data collection and release dates, the V2023 estimates had to rely on the 2022 ACS to estimate foreign-born immigration for 2022–2024 (holding the 2022 total constant over the subsequent years). The V2024 estimates use the 2023 ACS for 2023–24. The revision to these years is equal to the difference between foreign-born immigration totals in the 2023 ACS and the 2022 ACS.

On the second source of revision, we calculate annual humanitarian migration by region for estimate years 2021, 2022, and 2023 from same administrative data sources as the V2024 estimates. However, instead of using total humanitarian migration to inflate the ACS totals, we tabulate humanitarian migration by region of origin.<sup>34</sup> For July to December 2024, we follow the Census Bureau’s approach and use only the ACS estimate of foreign-born immigration.

We next distribute the annual revision to foreign-born immigration to demographic characteristics. Unlike the Census Bureau, which produces population estimates for single years of age, sex, race, and Hispanic origin, we need only to match the demographic detail used for the CPS second-stage weights. We construct proxy universes for each year and region of foreign-born individuals who arrived in the United States within the past five years tabulated from pooled 3-year ACS samples. For each CPS population target (national coverage step, second-stage race/sex/age, and second-stage eth-

---

34. We distribute admittances that were not from one of the 10 countries reported in Department of Homeland Security, Office of Homeland Security Statistics (2025) to regions using the regional share of total encounters, by agency, in a quarter.

nicity/sex/age), we calculate the share of each demographic group in the total population, multiply the inflows for each region by the demographic group shares, and sum across all regions.

To construct our alternative population estimates, we take the annual humanitarian migration totals by population target and demographic group and calculate the monthly inflow. We follow the V2024 methodology and assume that the inflows occur evenly over a year. For example,  $\frac{1}{12}$  of the annual total for estimate year 2023 would be added to each month from August 2023 through July 2024. Finally, for each population target we take the V2023 estimates and add the cumulative sum of the monthly inflows.