WERN Workers' Survey Raking Weights Appendix

5/25/23

We relied on the 2021 American Community Survey downloaded from IPUMS USA (Ruggles et al. 2023) to construct two sets of sampling weights (`rk_wgt_ind`, and `rk_wgt_final`) using raking. The ACS is nationally representative survey with 1-in-100 random sample of the population. The ACS data provides us with approximations to the marginal distributions of key demographic groups in the population of interest (people employed in one of our five targeted industries in 2022). Based on the available ACS data and the survey sampling process, we use the following variables to create weights: gender (female, not female), race (white, black, Hispanic/Latinx, and other[1]), age (18-24, 35-54, 55+), and industry (healthcare, hospitality, retail, telecommunications, warehousing).

To construct the industry-level marginal distributions, we map the NAICS industry codes of ACS workers' primary jobs into the five industries sampled in WERN workers' survey:

WERN survey industries	**NAICS**
Healthcare	621, 622, 623
Hospitality	71, 72
Retail	44, 45, 4MS (other retail trade)
Telecommunications	517
Warehousing	493

We created two weights 1) industry-specific weights and 2) final weights.

1. Industry-specific weights (`rk_wgt_ind`) are created by raking within each of the 5 industries, calculated using industry-specific gender, age, and race distributions for employed workers from the 2021 ACS.
2. Final weights (`rk_wgt_final`) are calculated by re-weighting the industry-specific weights by the share of workers in the five industries from the 2021 ACS.

To calculate raking weights, we used the `pewmethods` R package. Below are the steps to generate the weights specified above.

1. We recode gender, race, age, and industry variables from ACS to reflect WERN's sample.
    a. Note that we do not include self-identified white Hispanics as part of our Hispanic/Latinx racial group. We also recode black Hispanics as black.
2. We first create industry-specific weights by subsetting the sample to each of the industry, and rake weights based on three targets (age, gender, race). Once the process is complete, we merge each subsetted sample into one.
3. We then create the final weights by:
    a. Create cross-industry weights where each of the five industry groups are weighted based on the that industry's share of employed workers across all five industries (i.e., not considering other industries) in the ACS.
    b. We then multiply the industry-specific weights and the cross-industry weights.

[1] This includes those who identified as Asian/Pacific Islander, Native American and "other."

c. Finally, we trim the weights at the 1st and 99th quantiles to arrive at the final weight. We trim weights to reduce the design effect from weighting (according to Pew Research).

Reference:

Steven Ruggles, Sarah Flood, Matthew Sobek, Danika Brockman, Grace Cooper, Stephanie Richards, and Megan Schouweiler. IPUMS USA: Version 13.0 [dataset]. Minneapolis, MN: IPUMS, 2023. https://doi.org/10.18128/D010.V13.0