Propensity-Adjusted Raking with Applications to Current Population Survey Nonresponse

Justin J. McIllece[1]
[1]U.S. Bureau of Labor Statistics, 2 Massachusetts Ave NE, Washington, DC 20212

Abstract
Household nonresponse in the U.S. Current Population Survey (CPS) is addressed through a single ratio adjustment at cell levels defined by geography and urbanicity. Most nonresponse cells are defined within state boundaries, combining primary sampling units that have similar metropolitan status and population size. Though the CPS is a panel survey with differential response rates by rotation group and county level, the current nonresponse adjustment procedure accounts for neither due to instability concerns arising from small adjustment cells. In this paper, Bayesian response propensity adjustment factors (RPAFs) are constructed at the county level, leveraging longitudinal relative response rates to inform a preliminary weight adjustment. The single nonresponse vector is then expanded into a two-dimensional matrix, incorporating rotation group, before raking the RPAF-adjusted sample weights to the marginal totals.

Key Words: Current Population Survey, CPS, nonresponse, raking, response propensity

1. Introduction

The Current Population Survey has benefitted from sterling response rates throughout its history, but no survey is immune from the global rise in nonresponse afflicting the field.

A monthly household survey, the CPS has measured the state of the American labor force for about 80 years, achieving an average unit response rate of 93 percent as recently as the mid-to-late 1990s[2]. But the turn of the century marked the beginning of the decline—almost imperceptibly at first—as complex forces applied downward pressure on response propensities throughout the world[3].

Figure 1 displays the recent acceleration of CPS nonresponse, from a gradual loss of one or two percentage points in the 2000s; to the ten-point decline across the 2010s; through the Covid-19 data collection disruptions of 2020 to the recent but tenuous 2023 stasis around 70 percent.

[1] Views expressed are those of the author and do not necessarily reflect the views or policies of the U.S. Bureau of Labor Statistics.
[2] Computed from CPS microdata, January 1994–December 1999.
[3] The reasons behind declining response rates are not the focus of this paper. For discussion of CPS nonsampling error, see *Current Population Survey Design and Methodology, Technical Paper 77* (TP77; Census 2019).

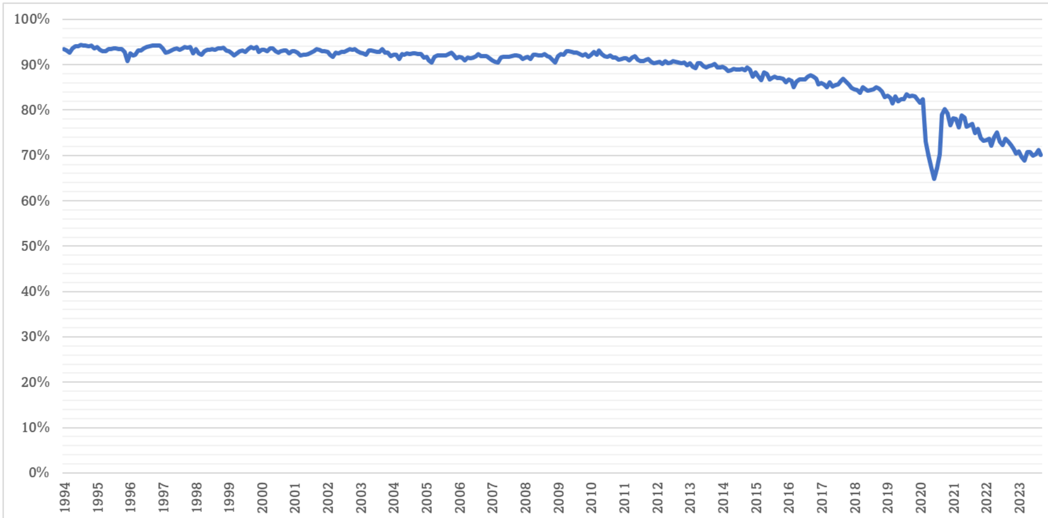

Figure 1: Unweighted CPS household response rates, January 1994–September 2023.

Current nonresponse adjustment procedures (§2) are essentially unchanged from two decades ago, when the nonresponse rate hovered around seven percent nationally; however, in the first nine months of 2023, the average nonresponse rate was about thirty percent, or 4.5 times higher. Thus, the CPS nonresponse adjustment factors (NRAFs) are doing nearly five times as much corrective work as they were in 2003. In the current environment, unrepresentativeness propagated throughout the nonresponse adjustment phase of weighting has much greater potential to bias labor force estimates at various demographic levels of detail. Overt or covert deficiencies in NRAF computation are now magnified.

Nonresponse adjustment in the CPS utilizes single-cell weighting adjustments, in which the sample weights of all responding households within a given cell, or cluster, are increased by the same adjustment factor. The clusters themselves are constructed geographically, designed to include enough monthly respondents to ensure stability of the adjustment factors and avoid variance inflation. When response rate decline accelerates, the bias reduction achieved by these cells is of increasing importance, but eventually they must expand to guarantee sufficient respondent counts. The larger nonresponse cells grow, the more their ability to counteract nonresponse bias is attenuated.

The Gordian Knot (Andrews, 2023) of modifying nonresponse adjustment procedures as a remedy against declining response rates—i.e., to do more with fewer respondents—is an obvious snarl to successfully mitigating biases associated with (or exacerbated by) said decline. Reconfiguring expanding cells to reduce potentially increasing nonresponse bias is akin to untangling the impossible knot, whereas the Alexandrian solution of slicing a sword through it might be analogized to reframing the problem itself: to do more with *more* respondents.

In this paper, an overview of pertinent CPS sampling and weighting methods is presented (§2); Bayesian response propensities are computed to inform a novel pre-adjustment for nonresponse (§3); current one-dimensional nonresponse adjustment is expanded into two dimensions (§4); and lastly, some conclusions are drawn with recommendations for continuing research (§5).

2. CPS Sampling & Weighting

The CPS sample is "redesigned" once every ten years. Redesign comprises redefinition and stratification of primary sampling units (PSUs), allocation and selection of new and continuing

PSUs within each state, and ultimately the 16-month phase-in/phase-out period[4] as the previous sample is transitioned out and the new sample is rotated in (TP77).

Crucial to the context of this paper, the selection of PSUs—either counties or groups of contiguous counties—remains in place ten years before either phasing out of or continuing into the succeeding sample. The majority of sampled PSUs are self-representing metropolitan areas, meaning their selection probabilities are equal to one. The overlap of self-representing (SR) PSUs from one decadal sample to the next is extremely high, as there are typically few if any major changes to the SR definitions (Nguyen and Gerstein, 2011). For non-self-representing (NSR) PSUs, CPS sampling methods average about 60-percent overlap (Rottach and Murphy, 2009), as the sample redesign has minimum continuity requirements for practical reasons related to budget and in-person data collection[5].

The CPS combines counties into nonresponse adjustment clusters based on metropolitan status, which is presumed to be a useful geographic proxy for response propensities and labor force conditions[6]:

1. Metropolitan, principal city
2. Metropolitan, not principal city
3. Nonmetropolitan (or rural)

Within these three categories, counties of similar size are combined such that the resultant nonresponse adjustment clusters are expected to be sufficiently large throughout the life of the corresponding sample[7]. The sample sizes in many county by metropolitan status subdomains are simply too small to apply adjustment factors directly. For example, the rural area of a mostly urban county might have zero or single-digit sample households in a given month.

After nonresponse adjustment, a series of weighting steps are conducted:

- First-stage weighting
- National coverage step
- State coverage step
- Second-stage weighting
- Composite estimation

Unlike nonresponse adjustment, these five weighting steps all incorporate external population estimates. First-stage weighting and the national and state coverage steps are single ratio adjustments designed to improve coverage.

The final two steps are iterative methods. Second-stage weighting further improves coverage, benchmarking to population controls by

- State/sex/age
- Ethnicity/sex/age
- Race/sex/age

---

[4] The CPS utilizes a 4-8-4 rotation design. Households are in sample four months, out of sample the next eight months, then back in sample four additional months.

[5] Data is collected via personal visit interviews for the predominance of first month-in-sample and fifth month-in-sample households.

[6] Urban areas tend to have lower response rates than rural areas, and labor force conditions between urban and rural areas within a state often exhibit differences.

[7] The target minimum respondent size for CPS nonresponse adjustment clusters is 50 households. Occasionally, a few cells slip below this threshold, but the vast majority exceed it.

The level of demographic detail in each of these categories varies by dimension. For example, the national ethnicity and race dimensions allow for greater sex by age detail, while the state dimension is more limited.

Finally, composite estimation improves the precision of national estimates of over-the-month change, particularly for stable labor force conditions such as employment[8]. Unlike prior steps, compositing does not have a coverage benefit and induces some bias into labor force estimates (Erkens 2012, 2017).

2.1 CPS Weighting Modifications

Cross-sectionally, sample size and response rate practicalities prevent any meaningful reconfiguration of the nonresponse adjustment clusters, especially considering that long-term consistency of the labor force time series is imperative.

Historically, modifications to CPS weighting procedures typically arose from expanded racial definitions and/or the desire to improve the precision of some facet of the estimator.

Though not comprehensive, a few of these changes (TP77) were:

- In 1953, racial definitions were first included in CPS weighting procedures, allowing for more precision of labor force estimates by White and non-White populations.
- In 1968, racial definitions were expanded to White, Black, and Other. Other (using modern categories) comprised all non-White and non-Black persons, such as Asian; American Indian and Alaska Native; and Native Hawaiian or Pacific Islander.
- In 1979, second-stage weighting was modified to improve the reliability of metropolitan and nonmetropolitan labor force estimates.
- In 1985, Hispanic ethnicity was added as a population control, dramatically improving precision of Hispanic labor force estimates.
- In 1998, composite estimation was first implemented in the CPS, placing weight on both past and current months' second-stage estimates to reduce the variance of over-the-month and over-the-year change.
- In 2003, racial definitions were expanded to the 31 categories still in use as of October 2023, consistent with the 2000 decennial Census, including information on mixed-race persons. The cellular detail of composite weighting was also expanded considerably to its current form.

These illustrative examples focus on weighting modifications, as that is the focus of this paper, but omit many affective changes to sampling procedures and external population controls (to which CPS weights are benchmarked) as well as the introduction and evolution of seasonal adjustment techniques to labor force estimates.

All are designed to improve coverage, representativeness, and precision via calibration and compositing, but none are direct modifications to the nonresponse adjustment step. This is hardly surprising, given the exceptional response rates noted in §1. Statistically interpreting this abbreviated history, throughout its first 60 years of existence, the CPS implemented substantial weighting changes primarily to compute better level and change estimates for an increasing number of racial and ethnic groups. The primary means of facilitating these changes were second-stage weighting and, between 1998 and 2003, composite estimation. The weighting modifications were responses to the demands of the times.

Recent official research has not unearthed major systematic biases in primary labor force estimates as a result of decreasing response rates or the Covid-19 pandemic (McIllece 2020; Rothbaum and

---

[8] Over-the-month change estimates of unemployment gain little precision from compositing due to lower month-to-month correlations.

Bee 2021[9]), and the meta-analysis of response rates and estimation bias in surveys by Groves and Peytcheva (2008) suggests that surveys with response rates around 70 percent are somewhat less likely to suffer from major nonresponse bias than surveys with lesser response rates[10]. The CPS *might* be reasonably protected against labor force bias if able to maintain 2023 response rates into the future, but it would be foolish to assume so. Over the past 20 years, the response rate decline (Figure 1) indicates the need, at minimum, for some review of nonresponse adjustment procedures and research into enhancements or alternatives.

3. Response Propensity Adjustment Factors

Considering that the selected counties remain in the CPS sample for at least ten years, the idea from §1 to "do more with more respondents" can be viewed as a Bayesian solution to the response rate problem. Rather than compute static monthly NRAFs, longitudinal response propensity adjustment factors (RPAFs) can be utilized as an alternative, leveraging the history of CPS sample and respondent behavior in a particular area to inform an appropriate adjustment factor.

The latter approach is strengthened by pooling past data to increase the effective sample size, not only mitigating the effects of declining response rates on the current nonresponse adjustment clusters but also enabling the disaggregation of them. Rather than combining multiple county by urbanicity areas in the longitudinal method—a necessity for independent NRAFs, despite response rate or measurement differentials—each detailed cluster can remain distinct if tethered to a reasonable prior for response propensity.

The natural choice, which also maintains as much consistency as possible with current methodology, is the overall response propensity from its existing nonresponse adjustment cluster. Thus, for a newly sampled area, its nonresponse adjustment will most heavily depend on the overall response tendency of its usual cluster. As time goes by and the effective sample size increases, additional information can be combined with the prior to compute a posterior estimate of response propensity. The RPAF is simply the inverse of that posterior estimate.

This method offers some attractive properties, fitting in neatly with current procedures rather than overhauling them (important for maintenance of time series) and only creating impactful adjustments in nonresponse clusters that need them. For instance, if a nonresponse cluster aggregates several detailed areas, and those areas all have similar response propensities, then the result of applying the Bayesian RPAF would be approximately the same as the static NRAF in production. Lastly, though the RPAFs are computed at a finer geographic detail, their behavior tends to be smoother overall, ameliorating concerns about variance inflation (§3.2.3).

3.1 Longitudinal Response Rates

Response propensity models are a common alternative to (or a definitional input into) traditional weighting class adjustments. These models can take various forms, such as logistic or probit regression, among other methods. Little (1986) explains the risk in direct application of inverse, modeled propensities as nonresponse-adjusted weights, warning that "respondents with very low (estimated response propensities) receive large weights that can inflate the variance of survey estimates excessively." Little continues: "Another argument for (response propensity) stratification is that it places less reliance on correct specification of the response propensity…since the

---

[9] Using linked administrative data, Rothbaum and Bee (2021) concluded that higher-income households were more likely to respond to the CPS Annual Social and Economic Supplement during the pandemic, but they did not associate this with bias in monthly labor force statistics like the unemployment rate.

[10] Groves and Peytcheva (2008) stress that "conclusions must be made with considerable caution" and, in the strongest of their four conclusions, state: "High response rates can reduce the risk of bias. They do this less when the causes of participation are highly correlated with the survey variables."

predictions are used only to partially order the sample, rather than to supply probabilities to be directly used in weighting."

A weighting class adjustment, then, might be viewed as a special case of general response propensity stratification, in which the implicit model is simply a uniform response likelihood for all sample units within a cluster, and differing response propensities are assumed between the clusters.

The approach adopted here is most similar to traditional weighting class adjustment, the difference being that instead of assuming a response propensity for all sample units equal to the nonresponse cluster's monthly response rate, the response propensity is based on *longitudinal* response rates (LRRs), or accumulated response rates over time, exploiting the long-term inclusion of each CPS sample area.

Defining $\pi_{j,T}$ as the LRR of current nonresponse cluster *j* in estimation month *T* and $\pi_{k\in j,T}$ as the LRR of county *k* within cluster *j* in estimation month *T*:

$$\pi_{j,T} = \frac{\sum_{t=1}^{T} r_{j,t}}{\sum_{t=1}^{T} n_{j,t}}$$

$$\pi_{k\in j,T} = \frac{\sum_{t=1}^{T} r_{k\in j,t}}{\sum_{t=1}^{T} n_{k\in j,t}}$$

where

$r_{j,t}$ = number of responding households in nonresponse cluster *j* in month $t \in [1, \ldots, T]$

$n_{j,t}$ = number of sampled households in *j* in month *t*

$r_{k\in j,t}$ = number of responding households in county *k* within nonresponse cluster *j* in month *t*

$n_{k\in j,t}$ = number of sampled households in $k \in j$ in month *t*

An apparent contradiction to the conditions that initiated this research (§1) is that declining response rates over time would mean that longitudinal response rates would be biased and unreflective of true response propensities in month *t*. If a county were newly selected in the 2010 sample design, it would first enter data collection in April 2014 based on the CPS phase-in/phase-out design (TP77) and would remain in sample at least until the 2020 sample design became fully integrated a decade later. As shown in Figure 1, response rates in 2014 or 2015 were considerably higher than response rates in 2021 or 2022. An LRR computed through month *T* would almost certainly overstate response propensity, as it would include prior years' response rate information[11].

A seamless solution to this problem is to compute LRRs at the component level *relative* to the LRR of its nonresponse cluster. Under this construction, the assumption is that the relative response propensity of a detailed sample area to its current nonresponse cluster is constant[12]. The current cluster will contain the information about declining (or other dynamic) response rates; within the detailed area, in this case county by urbanicity, only the relative propensity is required.

To compute the relative propensity, first the expectation of $\pi_{k\in j,T}$ is calculated as the average $\pi_{j,t}$ over the time period from $t = 1, \ldots, T$, weighted by $n_{k\in j,t}$, the monthly sample sizes of county *k*

---

[11] This problem could easily impact other types of models, as well. For instance, a logistic regression model would need to be restricted to time periods with comparable responsiveness, which limits model observations and can be difficult to ascertain.

[12] There may be a "fully Bayesian" construction that obviates this assumption, but that extension is not considered in this paper.

within nonresponse cluster j. Weighting by the sample counts in k is critical to accurately reflect both the changing sample composition (such as during the phase-in/phase-out of newly sampled areas) and the expected response propensity of any county within the cluster. For example, if one county had a heavier sample density earlier in the timeframe, it would have a higher expected LRR than a county with more sample later in the reference period.

$$E\left(\pi_{k\in j,T}\right) = \frac{\sum_{t=1}^{T}\left(\pi_{j,t} n_{k\in j,t}\right)}{\sum_{t=1}^{T} n_{k\in j,t}} = \text{expected LRR of county } k \text{ in cluster } j \text{ in month } T$$

$$\hat{f}_{k\in j,T} = \frac{\pi_{k\in j,T}}{E\left(\pi_{k\in j,T}\right)} = \text{relative LRR of county } k \text{ in cluster } j \text{ in month } T \qquad (1)$$

The ratio $\hat{f}_{k\in j,T}$ can then be interpreted as an estimate of relative response propensity of county k within nonresponse cluster j in month T.

3.2 Bayesian Response Propensities

The distribution of the $\hat{f}_{k\in j,T}$ values with a minimum cumulative sample size of 100 households (for stability) is plotted in Figure 2:

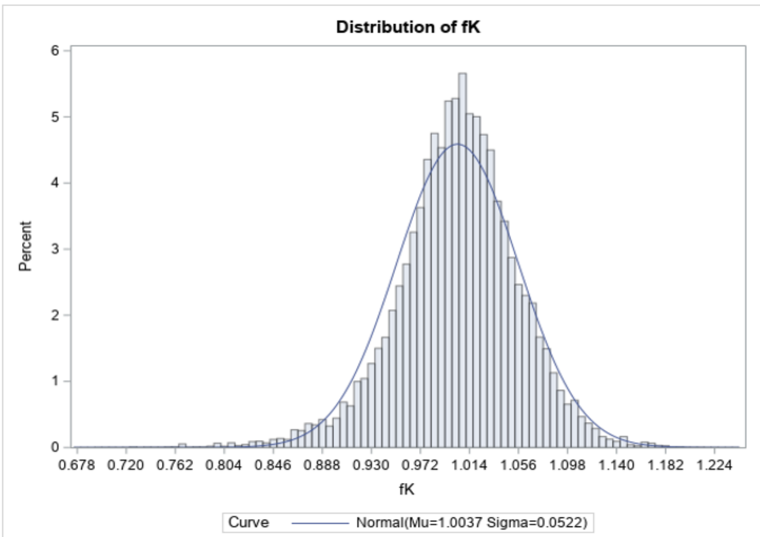

Figure 2: Histogram of national $\hat{f}_{k\in j,T}$ (relative response propensity) ratios from April 2014–Dec 2022. Minimum cumulative sample size of 100 households.

While diagnostic tests suggest the distribution is not quite Normal, in this paper it is treated as Normal for purposes of estimating the prior distribution of the relative response propensities. It is important that the mean of the prior distribution can be set to 1.00, which would indicate that any component county is initially expected to have equal response propensity as that of its membership nonresponse cluster. The estimated parameter of 1.0037 in Figure 2 suggests this assumption is reasonable.

Thus, the prior distribution for the relative response propensity of county k in cluster j is estimated to be

$$\tilde{f}_{k\in j} \sim N(1.00, 0.0027)\ \forall\ t \in 1, \cdots, T \qquad (2)$$

where $0.0027 = 0.0522^2$ is the estimated variance.

At time T, the observed relative response propensities are also assumed to be Normally distributed with mean relative response propensity ratio $\hat{f}_{k\in j,T}$ computed directly from the data using equation (1).

To estimate the variance of the observed ratios, since the response rate components are proportions, the binomial approximation to a Normal distribution is applied, treating $E(\pi_{k\in j,T})$ in the denominator of (1) as a constant for computational simplicity[13]. This results in the following variance approximation for the observed data in (1):

$$V(\hat{f}_{k\in j,T}) = V\left[\frac{\pi_{k\in j,T}}{E(\pi_{k\in j,T})}\right] \cong \frac{V(\pi_{k\in j,T})}{E(\pi_{k\in j,T})^2} = \left[\frac{1}{E(\pi_{k\in j,T})^2} * \frac{\pi_{k\in j,T}(1-\pi_{k\in j,T})}{\sum_{t=1}^{T} n_{k\in j,t}}\right] \tag{3}$$

Another consideration for equation (3) is the sample size component of the binomial approximation. Due to the rotating panel design, households in a given month T may have been included up to eight times in the denominator, meaning that the observations are not all independent. It may be more reasonable to use effective sample size rather than pure sample size in this calculation, which would result in a larger variance estimate for the data distribution of $f_{k\in j,T}$. However, the results presented in this section are based on pure household counts rather than effective sample size; the latter modification is left to future research (§5).

Given the approximate variance in (3), the data distribution is estimated as:

$$f_{k\in j,T} \sim N\left(\hat{f}_{k\in j,T}, \left[\frac{1}{E(\pi_{k\in j,T})^2} * \frac{\pi_{k\in j,T}(1-\pi_{k\in j,T})}{\sum_{t=1}^{T} n_{k\in j,t}}\right]\right)$$

Lastly, the posterior mean of two Normal distributions is calculated as a weighted average of the prior mean and observed mean, where the weights are the inverses of the respective variances (Murphy 2007).

If the weight for the observed $\hat{f}_{k\in j,T}$ is defined as $w_{k\in j,T}$, then:

$$w_{k\in j,T} = \frac{0.0027}{0.0027 + \left[\frac{1}{E(\pi_{k\in j,T})^2} * \frac{\pi_{k\in j,T}(1-\pi_{k\in j,T})}{\sum_{t=1}^{T} n_{k\in j,t}}\right]}$$

and the weight on the prior $\tilde{f}_{k\in j}$ is equal to $(1 - w_{k\in j,T})$.

Recall that the mean of the prior distribution is equal to 1.00. Thus, the posterior mean $f^*_{k\in j,T}$ can be reduced to:

$$f^*_{k\in j,T} = (1 - w_{k\in j,T}) * 1 + w_{k\in j,T} * \hat{f}_{k\in j,T}$$

$$f^*_{k\in j,T} = 1 - w_{k\in j,T} + w_{k\in j,T} * \hat{f}_{k\in j,T}$$

$$f^*_{k\in j,T} = 1 + w_{k\in j,T} * (\hat{f}_{k\in j,T} - 1) \tag{4}$$

Because the factor in (4) is the posterior estimate of relative response propensity, the inverse of (4) becomes the RPAF of county k relative to nonresponse cluster j, calculated over the time period t =

[13] Alternatively, Taylor Series linearization or replication could be applied to the ratio in (1) to better estimate the data-based variance.

1, …, T. In month T, for each k, the RPAF is multiplied by the monthly NRAF for the overall cluster, thereby accounting for response rate and labor force differentials among the member counties[14].

3.2.1 Constrained Response Propensity Adjustment Factors

The RPAFs computed in §3.2 are uncontrolled. This can be problematic when the county composition of the nonresponse cluster changes, which commonly occurs during the phase-in/phase-out of the decadal sample. In that 16-month transitional period, in the non-self-representing PSUs, some sample areas are phasing out of the sample while new areas are phasing in. As a result, the relative response propensities of continuing counties may be affected by the transition; i.e., if one or more new counties enters the nonresponse cluster with different response propensities than those they replaced, the relative propensities of the continuing counties will shift.

One option is to avoid the problem entirely by restarting the response propensity computations at the beginning of the phase-in of a new sample design. In this approach, the cumulative sample of both the counties and the nonresponse cells are reset to zero before the first month of the phase-in. This would be appropriate if the probabilistic samples from one design to the next were completely independent with low overlap.

However, in the CPS, the sample overlap is about 80 percent, as discussed in §1. In this setting, much information loss would be incurred by resetting the cumulative sample sizes to zero, and the loss of consistency in the weight adjustments would seem a heavy price to pay to sidestep a minor impediment.

As a compromise between free, continuing estimation of the RPAFs and restarting them with the phase-in of the new sample, a constrained RPAF is computed to mitigate the issues associated with each of the above:

$$\left(\frac{\sum_{i=1}^{r_1} w_{i,1,T}}{f_{1,T}^{**}}\right) + \left(\frac{\sum_{i=1}^{r_2} w_{i,2,T}}{f_{2,T}^{**}}\right) + \cdots + \left(\frac{\sum_{i=1}^{r_K} w_{i,K,T}}{f_{K,T}^{**}}\right) = \sum_{i=1}^{r_j} w_{i,j,T}$$

where r_k is the household response count in county $k = 1, \ldots, K$ in nonresponse cluster j, and r_j is the total household response count in cluster j.

Here, the inverse response propensities are modified such that the sum of the RPAF-adjusted, weighted response counts of the K member counties is equal to the weighted response count of the nonresponse cluster[15], which implies the simple ratio adjustment to the free response propensities computed in (4):

$$f_{k,T}^{**} = f_{k,T}^{*}\left[\frac{\sum_{i=1}^{r_j} w_{i,j,T}}{\sum_{k=1}^{K}\left(\frac{\sum_{i=1}^{r_k} w_{i,k,T}}{f_{k,T}^{*}}\right)}\right] \tag{5}$$

Controlling the posterior estimates in (5) corrects some (but not all) of the compositional shift affecting (4) while still leveraging long-term response history to inform the RPAF calculations. Because the RPAFs resulting from (5) are controlled to the nonresponse cluster, the NRAFs are

[14] The hierarchical nature of this adjustment method is extensible, such that an RPAF for the current nonresponse cluster j would be computed as a Bayesian posterior estimate over the same time period relative to, for instance, the statewide response propensity. Statewide propensities relative to national could also be considered. However, RPAFs were restricted to county level in this paper because it integrates most naturally with current nonresponse adjustment methods.

[15] The set notation from previous sections is dropped here for legibility.

unchanged[16], allowing for simple comparisons of representativeness with or without the RPAF adjustment (§3.2.2).

3.2.2 Case Study: New York City Demographics

As a motivating example, consider the five Boroughs of New York City. Each Borough is its own county, and all five counties are in the CPS sample[17]:

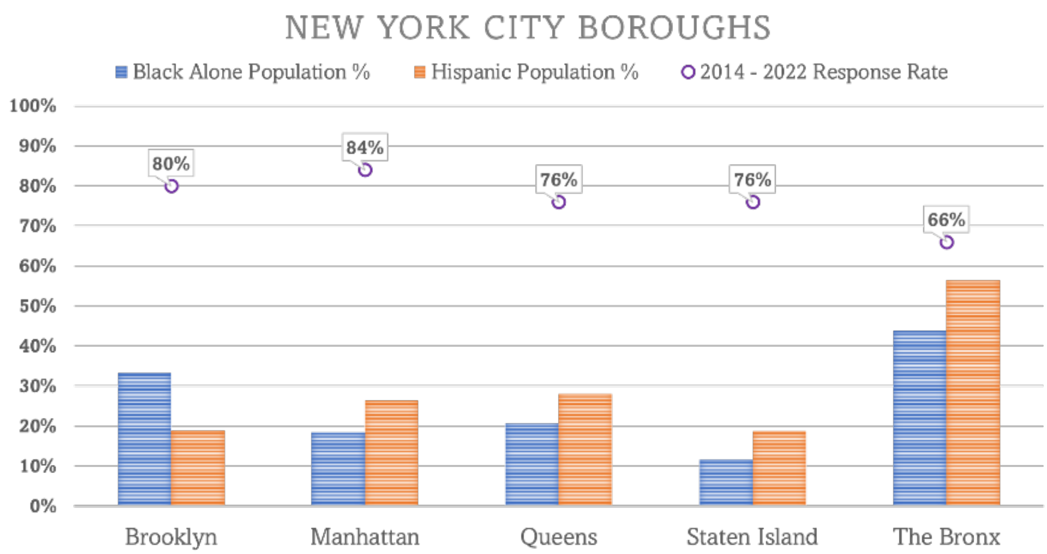

Figure 3: Longitudinal response rates and average Black alone and Hispanic percentages for the five Boroughs of New York City, January 2014–December 2022. Hispanic persons can be of any race.

The Black and Hispanic populations of New York City vary considerably by Borough and, critically, so do the response rates. The Bronx stands out for having the densest Black and Hispanic populations and the weakest response rates across the five Boroughs. According to demographic analysis conducted by Passel, Lopez, and Cohn (2022), the Bronx is one of a small number of counties outside of California and the Southwest with a majority Hispanic population. And while not shown in Figure 3, there are differences in labor force tendencies across the Boroughs, as well, thereby meeting the two well-known conditions for corrective nonresponse adjustment described by Little (1986) and more recently summarized by Kreuter and Olson (2011): "(E)ffective survey nonresponse adjustment variables should be highly correlated with both the propensity to respond…and the survey variables of interest."

Recalling the nonresponse adjustment methodology from §2, the five Boroughs are all combined into a single nonresponse cluster along with some other sample areas outside New York City. Therefore, each member county is upweighted by a single NRAF computed as the ratio of weighted sample households to weighted respondent households across the entire cluster, propagating the unrepresentativeness in Figure 3 into successive weighting adjustments.

Applying the constrained RPAFs from §3.2.1 to December 2022, the nonresponse cluster[18] is brought into better demographic balance:

---

[16] Subsequent steps, such as second-stage weighting and compositing, would be affected because they are person-level rather than household-level adjustments.

[17] While CPS sample information is confidential for privacy protection, the five Boroughs of New York City are all clearly identified as sample counties in the CPS public-use microdata files because publishing that information does not create a disclosure risk.

[18] Non-New York City counties in the nonresponse cluster are omitted from Table 1.

Table 1: Response Propensity Indicators in New York City
Selected Demographics, December 2022
(Based on January 2014–December 2022 data)

Borough	*Black alone* %	*Hispanic* %	*LRR* %	*RPAF*	*LRR * RPAF* %
Brooklyn	33	19	80	0.97	78
Manhattan	19	26	84	0.92	77
Queens	21	28	76	1.02	78
Staten Island	12	19	76	1.02	78
The Bronx	44	56	66	1.16	77

The complete nonresponse cluster has a 77-percent LRR over this timeframe. Of the five Boroughs, Queens and Staten Island (76 percent each) are most similar to the cluster as a whole, and their resulting RPAFs are close to one, indicating their base weights would be little changed due to relative response propensities. Brooklyn and Manhattan are overrepresented in the cluster and would be slightly downweighted by RPAFs of 0.97 and 0.92, respectively.

Most significantly, the Bronx is underrepresented to an uncomfortable degree, as evidenced by a 66-percent LRR, 11 percentage points below the total cluster. The RPAF method would inflate the weights of Bronx County respondents by the factor 1.16. In Table 1, comparing longitudinal response rates (column 4) to RPAF-adjusted response rates (column 6), the five Boroughs attain more balanced representativeness after accounting for county-level response propensities prior to the uniform nonresponse adjustment step.

Past research by Motel and Patten (2012) identifies Bronx County of New York as home to the largest populations of Dominicans and Puerto Ricans in the 50 states and the District of Columbia. While second-stage weighting benchmarks Black and Hispanic persons to statewide population controls (§2), it does not correct for within-state interactions between geography and demography.

Recently, the BLS began publishing monthly labor force estimates for Asian and Hispanic detailed ethnic groups (BLS, 2023), including Dominicans and Puerto Ricans, stressing the need to properly represent detailed labor force demographics as much as possible despite the sample size and response rate limitations of the CPS. Applying RPAFs at the county level could help in this regard[19], as demonstrated in the New York City example in Table 1.

3.2.3 National Representativeness and Efficiency

True of any weighting modification, stability is paramount. Precision of the estimator must not be artificially inflated. As a brief, though incomprehensive, evaluation of the efficiency of using Bayesian response propensities to inform a nonresponse pre-adjustment, national average RPAFs are computed for a few major racial and ethnic groups. Average coefficient of variation (CV) ratios are also calculated, relating the efficiency of the RPAFs to the current NRAF method. Both sets of metrics are presented in Table 2.

Table 2: Average RPAFs and CV Ratios
Selected Demographics, December 2022
(Based on January 2014–December 2022 data)

[19] Benchmarking to additional population controls in second-stage weighting is another possibility, either alongside or in place of modifications to nonresponse adjustment. The estimates that serve as population controls, subject to sources of error (Population Estimates Program, 2022), must be reliable and stable to improve efficiency of the estimates—not always the case with relatively small subpopulations.

Demographic	*Average RPAF*	*Average CV Ratio*
White	0.999	1.001
Black	1.003	1.001
Hispanic	1.001	0.999
Asian	1.001	1.000
AIAN (American Indian / Alaska Native)	1.001	1.003
NHPI (Native Hawaiian / Pacific Islander)	1.001	1.001

At the national level, there is little overall difference in representativeness or efficiency for the demographic groups included in Table 2. This is a nationally stable adjustment step that corrects for local imbalances, as shown in New York City (Table 1). Disruptions to major CPS time series seem unlikely but would require a more fulsome analysis to verify.

4. Two-Dimensional Nonresponse Adjustment

Besides the geographic component, there are well-known and -documented differences in response rates and labor force tendencies at the month-in-sample (MIS), or rotation group, level (Bailar, 1975; Breau and Ernst, 1983; Lent et al., 1994 and 1999; Erkens 2012 and 2017; McIllece 2020 and 2022).

MIS 1 represents a household's first month in sample, MIS 2 the second, etc., until MIS 8, its final month in sample[20]. Each MIS is designed to be a representative sample, and under ideal conditions, all eight would have equal response propensities and each would provide unbiased estimates of the labor force. In reality, there are clear and consistent differences across the eight MIS in both response and labor force tendency[21]. Prominently, MIS 1 typically has the lowest response rate (Figure 4) but the highest inclination of reporting employed or unemployed persons relative to the other rotation groups.

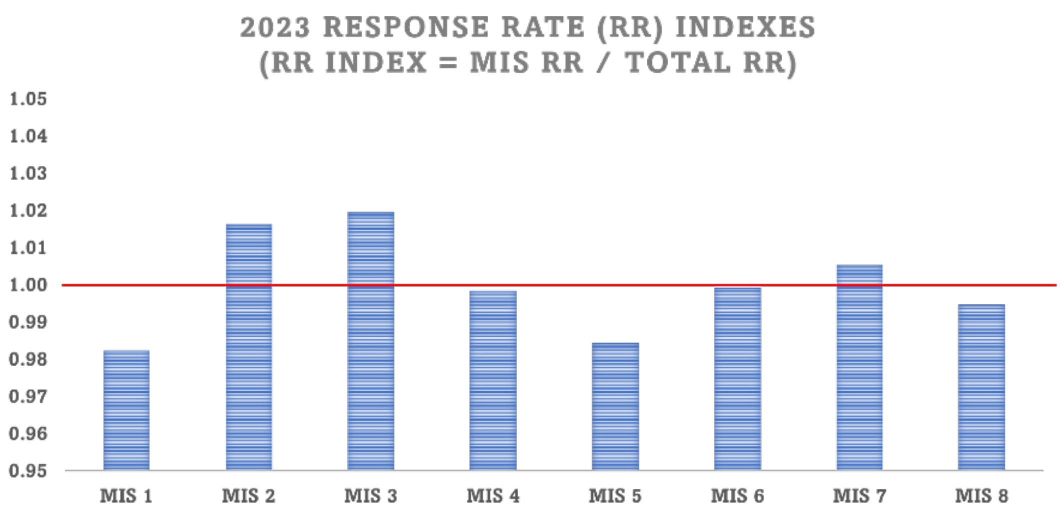

Figure 4: MIS response rate (RR) indexes, relative to total response rate. January 2023–June 2023 average.

[20] From the CPS 4-8-4 in-out-in rotation, MIS 5 occurs in the ninth month of the 16-month cycle, as a household then returns after being held out of the sample for eight months.
[21] There are likely various reasons for these disparities. A notable distinction is that most MIS 1 and MIS 5 interviews are conducted in person, while most MIS 2–4 and MIS 6–8 interviews are conducted by phone. MIS 1 and MIS 5, on average, have lower response rates and higher labor force tendencies.

Current CPS weighting procedures address MIS disparities in second-stage weighting, in which MIS are paired (1 and 5, 2 and 6, etc.) and benchmarked to external population controls (§2). However, considering the response rate differentials across the eight MIS, it makes intuitive sense to correct them individually in the nonresponse adjustment stage before pairing them for benchmarking.

4.1 Raking Coverage Rates

Empirically, MIS differentials appear to be a national phenomenon that cuts through all geographies and urbanicities, the categories that form the nonresponse adjustment clusters (§1). The lack of interactional evidence suggests a reasonable approach might be to apply standard CPS raking procedures to each state[22], where the current nonresponse clusters and the eight MIS are the two dimensions of iterative proportional fitting.

Self-evidently, if the current nonresponse clusters are sufficient for a one-dimensional NRAF, they are sufficient for the first dimension of a two-dimensional rake. The most recent CPS sample allocation guaranteed a minimum sample size of about 700 households per state (Rottach and Erkens, 2012); divided by eight, each state by MIS should comprise at least 85 households in any given month, on average. Assuming the 70-percent response rate observed in 2023 (Figure 1) would yield a minimum of about 60 responding households in the second dimension of the rake, exceeding the criterion established for the current nonresponse clusters[23].

The raking methodology in this extension is standard, the same as applied in CPS second-stage weighting, so it need not be restated here. Instead, summary results are presented below, highlighting potential benefits of expanding the nonresponse adjustment into two dimensions.

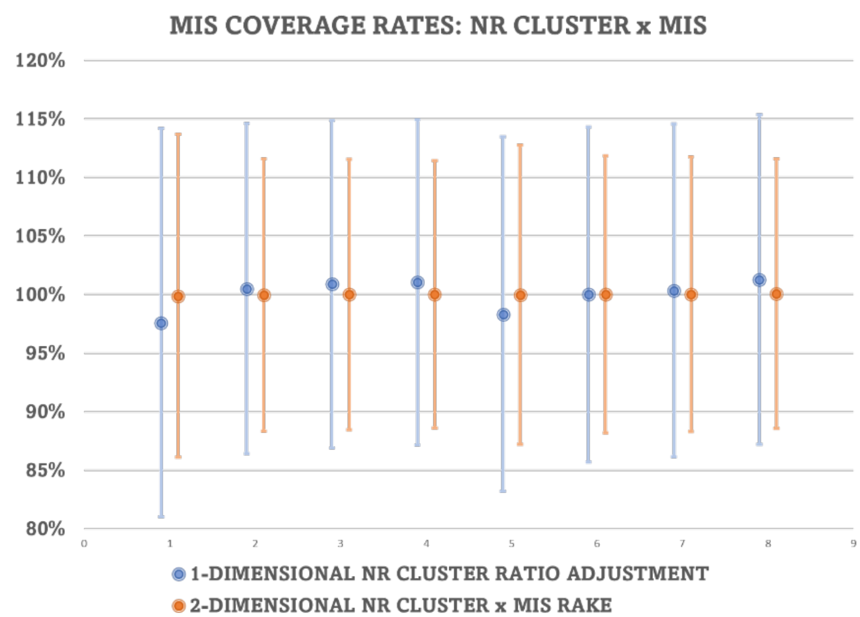

Figure 5: Comparison of 95-percent coverage intervals of the NR cluster by MIS cells using one-dimensional and two-dimensional nonresponse adjustment, May 2005–February 2020.

[22] For weighting purposes, CPS identifies 53 state or substates. Washington, DC, is treated as a state, while New York is separated into New York City and balance of state, and California is separated into Los Angeles County and balance of state.

[23] Due to variation in response propensities, some state by MIS cells would fall below 50 responding households.

As shown in Figure 5, expanding CPS nonresponse adjustment into two-dimensional raking improves coverage without appearing to destabilize the weights at this step. Presuming convergence, the post-nonresponse adjustment coverage of the raking procedure should be centered at 100 percent, as demonstrated here, whereas the coverage using one-dimensional NRAFs reflects the response rate index pattern observed in Figure 4. Additionally, the coverage distributions at the cluster by MIS level are tighter, indicating better representativeness of the cells in the two-dimensional grid.

4.2 Propensity-Adjusted Raking

Leveraging the benefits of the weighting modifications in sections §3.2.1 and §4.1, propensity-adjusted raking (PAR), as defined in this paper, combines the novelty of the Bayesian RPAFs with the foundational methods of CPS weighting. Because the two methods are independent[24], both the RPAF correction of local (county-based) response imbalances and the improved representativeness at the nonresponse cluster by MIS level should be retained.

In Figure 6, stacked histograms display county-level coverage rates using the CPS production NRAFs (top) and the PAR method (bottom):

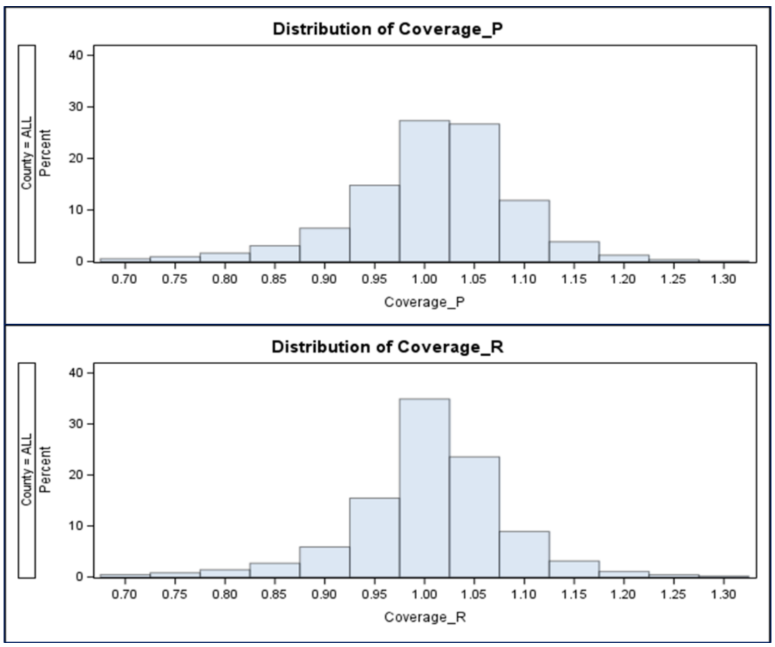

Figure 6: County coverage rates using current CPS nonresponse adjustment procedure (Coverage_P) and alternative PAR method (Coverage_R). May 2005–February 2020.

All counties nationally are included in the histograms. The PAR distribution has a substantially higher peak at 100-percent coverage, a smaller standard deviation, and less density in the tails. Overall, county representativeness is improved.

[24] In this context, independence means that the two weighting steps do not affect each other. They could be reversed in order and still produce the same results. Independence holds in this construction only when using the constrained RPAFs in equation (5).

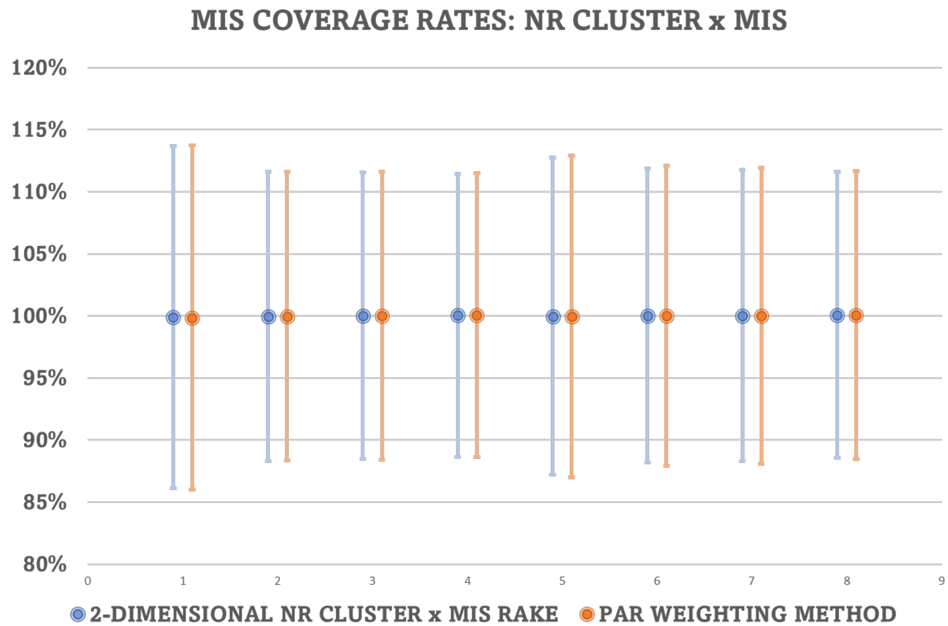

Figure 7: Comparison of 95-percent coverage intervals of the NR cluster by MIS cells using two-dimensional and PAR nonresponse adjustment, May 2005–February 2020.

Lastly, Figure 7 plots coverage rates at the same cluster by MIS level as Figure 5 but compares the two-dimensional rake to the hybrid PAR method. The two sets of coverage bars are almost identical, indicating that county-level RPAFs do not induce heightened variation in the nonresponse adjustment weights.

5. Conclusions

The golden age of CPS response rates ended around the turn of the century, as nonresponse gradually but inexorably accelerated throughout the 2000s and the 2010s before (excepting the temporal shock and recovery caused by the Covid-19 pandemic) reaching a recent equilibrium in the 2020s. Whether the consistence of 70-percent response rates in 2023 endures in the coming years is, of course, unknowable.

While there are other household surveys and data sources that contain direct or auxiliary information about the U.S. labor force, none offer both the quality and timeliness of the Current Population Survey. Thus, useful solutions to the problem of increasing CPS nonresponse must first come from within the survey itself.

In this research, longitudinal response rates inform Bayesian posterior estimates of county-level response propensities, relative to nonresponse clusters, thereby increasing the effective sample size for nonresponse adjustment despite nationally decreasing response rates. These response propensity adjustment factors are shown to be stable in their application while improving representativeness at the county level, an important step forward considering that the Bureau of Labor Statistics continues to publish more demographic labor force detail than in years past.

Nonresponse adjustment is expanded from one to two dimensions, incorporating household month-in-sample, a source of response rate and labor force differentials well known to CPS researchers of each decade back to the 1970s. Though understood for nearly half a century, MIS biases have been left for second-stage weighting to reconcile during that era. MIS bias can be efficiently addressed in nonresponse adjustment, improving coverage at the cluster by MIS level of cellular detail.

These two modifications are combined into propensity-adjusted raking for nonresponse, realizing both the benefits of better county representativeness—important for local geographies, such as the Bronx in New York City, home to the largest Puerto Rican and Dominican communities in the 50 states and Washington, D.C.—and improved coverage rates that account for MIS bias.

There are important limitations to the results presented in this paper, offering avenues for ongoing research:

The transition from one sample design to the next introduces changes to county composition in some nonresponse clusters, for which the Bayesian response propensities cannot perfectly account. The posterior estimates of relative response propensities may be biased by ignoring the repeated sampling of households in the rotation design. And county itself is not the end-all, be-all for correcting response rate differentials. Response propensities are likely more dependent upon socioeconomic characteristics; available on the sample frame, geographic information is used as a pragmatic proxy for such traits unknown about nonresponding households.

Thorough analysis of any weighting modifications requires more complete evaluation of the results than presented in this preliminary research summary. Changes must be fully tested by recreating the entire CPS weighting process resulting from the introduced methods, including all the post-nonresponse adjustment steps listed in §2. The bias and precision of labor force estimates—such as monthly levels, over-the-month changes, and statewide annual averages—must be compared to current estimation procedures for many important demographics, such as (minimally) those included in direct compositing. Avoiding substantial shifts or breaks in important time series is also critical, as the extensive, reliable history of CPS estimates is one of its core strengths.

Despite the noted limitations, given global response rate declines that have impacted virtually all official surveys, it seems incumbent upon CPS researchers to continue exploring potential improvements to the classical methods of nonresponse adjustment that have served the program well for many decades.

Acknowledgements

The author would like to thank BLS colleagues Nicholas Johnson, Stephen Miller, and Reid Rottach for helpful recommendations and review.

References

Andrews, E. (2023). "What Was the Gordian Knot?" *https://www.history.com/news/what-was-the-gordian-knot*. A&E Television Networks, LLC.

Bailar, B. (1975). "The Effect of Rotation Group Bias on Estimate from Panel Surveys," in *Journal of the American Statistical Association*, 70:349, pp. 23–30.

BLS Commissioner's Corner (2023). "Expanded Data for Detailed Hispanic or Latino Groups Now Available" and "BLS Now Publishing Monthly and Quarterly Labor Force Data for Detailed Asian Groups". *https://www.bls.gov/blog/archive.htm*. U.S. Bureau of Labor Statistics.

Breau, P. and Ernst, L. (1983). "Alternative Estimators to the Current Composite Estimator," in *Proceedings of the Section on Survey Methods Research, American Statistical Association*, pp. 397–402.

Current Population Survey Design and Methodology, Technical Paper 77, (2019). *https://www2.census.gov/programs-surveys/cps/methodology/CPS-Tech-Paper-77.pdf*. U.S. Census Bureau.

Erkens, G. (2012). "Changes in Panel Bias in the U.S. Current Population Survey and its Effects on Labor Force Estimates," in *Proceedings of the 2012 Joint Statistical Meetings, Section on Survey Research Methods*, pp. 4220–4232.

Erkens, G. (2017). "Practical and Theoretical Considerations of Panel Effects for the Current Population Survey's Composite Estimator," in *Proceedings of the 2017 Joint Statistical Meetings, Section on Government Statistics*, pp. 1449–1467.

Groves, R. and Peytcheva, E. (2008). "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis," in *The Public Opinion Quarterly*, 72:2, pp. 167–189.

Krauter, F., and Olson, K. (2011). "Multiple Auxiliary Variables in Nonresponse Adjustment," in *Sociological Methods & Research*, 40:2, pp. 311–332.

Lent, J., Miller, S., and Cantwell, P. (1994). "Composite Weights for the Current Population Survey," in *Proceedings of the Section on Survey Methods Research, American Statistical Association*, pp. 867–872.

Lent, J., Miller, S., Cantwell, P., and Duff, M. (1999). "Effects of Composite Weights on Some Estimate from the Current Population Survey," in *Journal of Official Statistics*, 15:3, pp. 431–448.

Little, R. (1986). "Survey Nonresponse Adjustments for Estimates of Means," in *International Statistical Review*, 54:2, pp. 139–157.

McIllece, J. (2020). "Covid-19 and the Current Population Survey: Response Rates and Estimation Bias," in *Proceedings of the 2020 Joint Statistical Meetings, Section on Survey Methods Research*.

McIllece, J. (2022). "Optimizing the Current Population Survey Composite Estimator," in *Proceedings of the 2022 Joint Statistical Meetings, Section on Survey Methods Research*, pp. 1974–1998.

Methodology for the United States Population Estimates: Vintage 2022 (2022). *https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2020-2022/methods-statement-v2022.pdf*. U.S. Census Bureau Population Estimates Program.

Motel, S., and Patten, E. (2012). "The 10 Largest Hispanic Origin Groups." https://www.pewresearch.org/hispanic/2012/06/27/the-10-largest-hispanic-origin-groups-characteristics-rankings-top-counties. Pew Research Center.

Murphy, K. (2007). "Conjugate Bayesian Analysis of the Gaussian Distribution." *https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf*. University of British Columbia.

Nguyen, T., and Gerstein, A. (2011). "Sample Design Research in the 2010 Sample Redesign," in *Proceedings of the 2011 Joint Statistical Meetings, Section on Survey Methods Research*, pp. 4359–4372.

Passel, J., Lopez, M., and Cohn, D. (2022). "U.S. Hispanic Population Continued Geographic Spread in the 2010s." *https://www.pewresearch.org/short-reads/2022/02/03/u-s-hispanic-population-continued-its-geographic-spread-in-the-2010s*. Pew Research Center.

Rothbaum, J., and Bee, A. (2021). "Coronavirus Infects Surveys, Too: Survey Nonresponse Bias and the Coronavirus Pandemic." in *Census Working Papers*. https://www.census.gov/library/working-papers/2020/demo/SEHSD-WP2020-10.html. U.S. Census Bureau.

Rottach, R., and Erkens, G. (2012). "Modeling Variances to Determine Sample Allocation for the Current Population Survey," in *Proceedings of the 2012 Joint Statistical Meetings, Section on Survey Methods Research*.

Rottach, R., and Murphy, P. (2009). "Maximizing Overlap of Large Primary Sampling Units in Repeated Sampling: A Comparison of Ernst's Method with Ohlsson's Method," in *Proceedings of the Federal Committee on Statistical Methodology Research Conference*.