Missing Data and Composite Estimators in the Current Population Survey

Morgan Heyde, Justin McIllece
Office of Employment and Unemployment Statistics, U.S. Bureau of Labor Statistics, 4600 Silver Hill Road, Washington, DC 20212-0002

Abstract

Until the U.S. government shutdown of October 2025, the Bureau of Labor Statistics (BLS) had consistently conducted the Current Population Survey (CPS) every month for decades. The CPS composite estimator, also known as an AK estimator, takes advantage of the 75 percent month-to-month sample overlap and reduces the standard error of over-the-month change. However, AK composite estimation relies on a measure of the change in overlapping panels which is not directly calculable for November 2025 since the prior month data is missing. In this paper, the alternative approach implemented in the November 2025 labor force estimates is developed, theoretical and empirical precision are evaluated, and the short-term and long-term impacts are considered.

Key Words: Current Population Survey, CPS, Composite Estimator, Missing Data, Variance Estimation, MIS effect

1. Introduction

The Current Population Survey is the source of the monthly household survey data for *The Employment Situation*, including the official U.S. unemployment rate. Due to a lapse in appropriations, the CPS did not collect data in October 2025. Since not seasonally adjusted CPS estimates are purely sample-based, no estimates could be produced, resulting in a permanent one-month gap in the historical time series. However, because the CPS utilizes a recursive composite estimator that relies on past months' estimates, this lapse also affected monthly estimates subsequent to October 2025.

Following the post-1950 Census, the CPS redesign introduced the 4-8-4 rotation system. Due to the typical 75 percent overlap of month-to-month samples inherent to this design change, implementing composite estimation reliably and significantly improved estimation by decreasing the standard errors associated with month-to-month changes (Lent 1999). This method was further updated in 1998 when BLS introduced an updated composite weighting methodology which determined better coefficients for various labor force categories by using ratio adjustments to fine-tune person weights (Census 2006).

2. The AK Composite Estimator

The composite estimation technique used by CPS is most often referred to as the AK composite estimator, which is a specialized case of the generalized composite estimator (GCE). This estimator relies on coefficient values A and K, where the A coefficient weights the term representing the difference between continuing months in sample and new months in sample. The K coefficient weights the average of the two estimators for the current month. Optimal A and K values for monthly labor estimates were determined through comprehensive testing and review as detailed in (Lent 1994).

$$\hat{Y}_t^{CMP} = (1-K)\hat{Y}_t^{SS} + K\left(\hat{Y}_{t-1}^{CMP} + \Delta_t\right) + A\hat{\beta}_t \tag{1.0}$$

$$\hat{Y}_t^{SS} = \sum_{i=1}^{8} x_{ti} \tag{1.1}$$

$$\Delta_t = \frac{4}{3}\left[\sum_{i=2}^{4}\left(x_{ti} - x_{t-1,i-1}\right) + \sum_{i=6}^{8}\left(x_{ti} - x_{t-1,i-1)}\right)\right] \tag{1.2}$$

$$\hat{\beta}_t = (x_{t1} + x_{t5}) - \frac{1}{3}\left(\sum_{i=2}^{4} x_{ti} + \sum_{i=6}^{8} x_{ti}\right) \tag{1.3}$$

Where

i = month-in-sample (MIS) 1, 2, ... , 8

x_{ti} = sum of second-stage weights of respondents in month t and MIS i with characteristic of interest

K = 0.4 for unemployed, 0.7 for employed

A = 0.3 for unemployed, 0.4 for employed

In the formulation above, the second term, $K\left(\hat{Y}_{t-1}^{CMP} + \Delta_t\right)$, relies on data from *t*-1, the prior month. $\hat{Y}_{t-1}^{CMP}$ is the estimate for the prior month and Δ_t is the difference term. This difference term quantifies the change in households that are in both the current and previous month. Note, if the prior month's data is missing, neither of these are directly calculable.

The first and last terms, $(1-K)\hat{Y}_t^{SS}$ and $A\hat{\beta}_t$ respectively, only require data for the non-overlapping MIS in the current month *t*. Thus, these terms can still be calculated in the event of missing *t*-1 data. Note that second-stage weights refer to CPS estimation weights after a series of poststratification, or benchmarking, adjustments to detailed demographic population controls produced by the Census Bureau (Technical Paper 77).

3. CPS 4-8-4 Sample Rotation Scheme

Month in Sample Pairs			
MIS	Previous	Current	Pair
1	→	R	
2	→	→	(1,2)
3	→	→	(2,3)
4	R	→	(3,4)
5	→	R	
6	→	→	(5,6)
7	→	→	(6,7)
8	R	→	(7,8)

Figure 1

MIS Pairs -Prior Month Missing				
MIS	Previous	Missing	Current	Pair
1	→	?		
2	→	?		
3	→	?	→	(1,3)
4		?	→	(2,4)
5	→	?		
6	→	?		
7	→	?	→	(5,7)
8		?	→	(6,8)

Figure 2

Figure 1 illustrates the month-in-sample pairs for the typical case where no data is missing. Note that this design results in 75 percent sample overlap from one month to the next since a household in MIS1 in the prior month will still be in the sample for the following month.

Figure 2 expands on Figure 1 by including a column to denote the missing month of data for October 2025. Since the AK composite estimator is designed to leverage the sample overlap, there are a few options to address the missing data column: data can either be left as is, which will result in a gap, or shifted to cover that gap. A gap would effectively ignore the October data and simply treat September 2025 data like September 2025 data (t-2), thus leaving October 2025 data (t-1) empty. However, this has larger theoretical and empirical bias than shifting. Shifting would simply shift every estimate before November 2025 up one month. The AK coefficients would correspond to the shifted MIS index, so in this example, the rotation group that was MIS1 in September 2025 will be treated as MIS2 in the shift composite, instead of MIS3 as it would in the usual case. This shift doesn't fully resolve the loss in overlap but still exploits the survey design, which results in better estimates than the gap method. Also note that this two-month change illustrated in Figure 2 results in only a 50 percent sample overlap, instead of the 75 percent shown in Figure 1.

4. Shifted AK Composite Estimator

The shifted AK composite estimator is defined as:

$$\hat{Y}_t^{CMP2} = (1-K)\hat{Y}_t^{SS} + K\left(\hat{Y}_{t-2}^{CMP} + \Delta_t^{(2)}\right) + A\hat{\beta}_t \tag{2.0}$$

$$\hat{Y}_t^{SS} = \sum_{i=1}^{8} x_{ti} \tag{2.1}$$

$$\Delta_t^{(2)} = \frac{4}{3}\left[\sum_{i=2}^{4}\left(x_{ti} - x_{t-2,i-1}\right) + \sum_{i=6}^{8}\left(x_{ti} - x_{t-2,i-1}\right)\right] \tag{2.2}$$

$$\hat{\beta}_t = (x_{t1} + x_{t5}) - \frac{1}{3}\left(\sum_{i=2}^{4} x_{ti} + \sum_{i=6}^{8} x_{ti}\right) \tag{2.3}$$

Where

i = MIS pair (i, j) where $i \in [1,2,3,4]$; $j = i + 4$

l = monthly lag

x_{ti} = sum of second-stage weights of respondents in month t and MIS i with characteristic of interest

K = 0.4 for unemployed, 0.7 for employed

A = 0.3 for unemployed, 0.4 for employed

When a month of data is missing, like October 2025, computing the AK composite estimator for the following month is impossible due to the incalculable terms, $\hat{Y}_{t-1}^{CMP}$ and Δ_t, which both rely on data from the missing month. The shift composite estimator remedies this by shifting the existing historical data forward to remove the gap, which shifts $\hat{Y}_{t-1}^{CMP}$ to $\hat{Y}_{t-2}^{CMP}$, and then by modifying the difference term Δ_t to reflect the two-month change in overlapping panels as $\Delta_t^{(2)}$. While this does result in a calculable composite estimator, using the two-month change only has 50 percent sample overlap where the usual AK composite estimator would have 75 percent.

5. Variance

To first compute the theoretical variance of the AK composite estimator, this paper follows the conventions of the variance calculations in McIllece (2022), which utilize the GCE form of the composite estimator and the following assumptions:

1. $V(x_{i,t}) = \sigma^2 \; \forall \; i, t$
2. $Cov(x_{i,t}, x_{j,t-l}) = r_{i,j,l}\sigma_t^2 \; \forall \; i, j, t, l$

Where the additional GCE terms are defined as

a_i = GCE month t composite coefficient for MIS pair i

b_i = GCE month t-l composite coefficient for MIS pair i

$d_i = \; a_i - \; b_i$

Using this form, Equation (3) below is an expansion of the covariance structure of the estimator, substituting the geometric identity $\sum_{t=0}^{\infty} ar^t = \frac{a}{1-r} \; \forall \; |r| < 1$ to produce a closed-form solution to the infinitely recursive AK composite estimator, assuming the lag is sufficiently large and the covariance matrix is static.

The resulting variance estimate for a monthly level AK composite estimate is given by

$$V(\hat{Y}_t^{CMP}) = \sum_{i=1}^{4}\sum_{j=1}^{4}\left(a_i a_j r_{i,j,0} + \frac{K^2}{1-K^2} d_i d_j r_{i,j,0}\right)\sigma^2 + 2\left[\sum_{l=1}^{L}\sum_{i=1}^{4}\sum_{j=1}^{4}\left(K^l a_i d_j r_{i,j,l} + \frac{K^2 K^l}{1-K^2} d_i d_j r_{i,j,l}\right)\right]\sigma^2 \quad (3)$$

Given the above assumptions, the variance of the shift composite estimator can be written as:

$$V(\hat{Y}_t^{CMP2}) = \sum_{i=1}^{4}\sum_{j=1}^{4}\left(a_i a_j r_{i,j,0} + \frac{K^2}{1-K^2} d_i d_j r_{i,j,0}\right)\sigma^2 + 2\left[\sum_{l=1}^{L}\sum_{i=1}^{4}\sum_{j=1}^{4}\left(\left(K^{l-1} a_i d_j r_{i,j,l}\right) I_l + \frac{K^2 K^l}{1-K^2} d_i d_j r_{i,j,l}\right)\right]\sigma^2 \quad (4)$$

Then the difference between the AK composite and the shift composite estimators is calculated as

$$V(\hat{Y}_t^{CMP}) - V(\hat{Y}_t^{CMP2}) = \sum_{i=1}^{4}\sum_{j=1}^{4}\left(a_i a_j r_{i,j,0} + \frac{K^2}{1-K^2} d_i d_j r_{i,j,0}\right)\sigma^2 + 2\left[\sum_{l=1}^{L}\sum_{i=1}^{4}\sum_{j=1}^{4}\left(K^l a_i d_j r_{i,j,l} + \frac{K^2 K^l}{1-K^2} d_i d_j r_{i,j,l}\right)\right]\sigma^2 - \sum_{i=1}^{4}\sum_{j=1}^{4}\left(a_i a_j r_{i,j,0} + \frac{K^2}{1-K^2} d_i d_j r_{i,j,0}\right)\sigma^2 + 2\left[\sum_{l=1}^{L}\sum_{i=1}^{4}\sum_{j=1}^{4}\left(\left(K^{l-1} a_i d_j r_{i,j,l}\right) I_l + \frac{K^2 K^l}{1-K^2} d_i d_j r_{i,j,l}\right)\right]\sigma^2$$

$$= 2\left[\sum_{l=1}^{L}\sum_{i=1}^{4}\sum_{j=1}^{4}\left(K^l a_i d_j r_{i,j,l} + \frac{K^2 K^l}{1-K^2} d_i d_j r_{i,j,l}\right)\right]\sigma^2 - 2\left[\sum_{l=1}^{L}\sum_{i=1}^{4}\sum_{j=1}^{4}\left(K^{l-1} a_i d_j r_{i,j,l} + \frac{K^2 K^l}{1-K^2} d_i d_j r_{i,j,l}\right)\right]\sigma^2$$

$$= 2\left[\sum_{l=1}^{L}\sum_{i=1}^{4}\sum_{j=1}^{4}\left(K^l a_i d_j r_{i,j,l}\right)\right]\sigma^2 - 2\left[\sum_{l=1}^{L}\sum_{i=1}^{4}\sum_{j=1}^{4}\left(K^{l-1} a_i d_j r_{i,j,l}\right)\right]\sigma^2$$

$$= 2\sum_{i=1}^{4}\sum_{j=1}^{4}\left(Ka_i d_j r_{i,j,1}\right)\sigma^2 - 2\left[\sum_{l=1}^{L}\sum_{i=1}^{4}\ \sum_{j=1}^{4}\left(K^{l-1}(K-1)a_i d_j r_{i,j,l}\right)\right]\sigma^2 \qquad (5)$$

The final Equation (5) can be used to estimate the expectation of the difference in the variances, which is particularly useful in generalized variance function standard error generation.

6. MIS Effects and Bias

MIS effects are a well-documented and -researched phenomenon wherein respondents' answers are systematically different depending on their current MIS group (Huang 1981, Erkens 2012). Given what is known about this phenomenon, there are two possibilities for the MIS effects following a missing month of data. The first case assumes that MIS2 respondents following a missing month will have the same MIS2 effect as any other MIS2, while the second case assumes that the MIS2 following a missing month will not have the same MIS2 effect as months with data for the previous month.

Since October data collection did not occur, the group labelled as month in sample 2, MIS2, was collected in the sample for the first time in November. Given the body of research on the MIS effect, it seems unlikely that this November 2025 MIS2 will respond as the first case suggests and instead, we would expect them to respond more like an MIS1.

Derivation omitted for brevity, the expected difference, or bias relative to the AK composite estimate, can be expressed as

$$E_2\left[\hat{Y}_{T+L}^{CMP2}\right] - E_0\left[\hat{Y}_{T+L}^{CMP}\right] = K^{L-3}[K^3\theta_1 + K^2\theta_2 + K\theta_3 + \theta_4]\ \forall\ L \geq 4 \qquad (6)$$

Where $\theta_1, \theta_2, \theta_3, \theta_4$ are constants which are defined by the form of the AK composite estimator and the MIS biases.

The important thing to note here is the K^{L-3} term which illustrates the exponential decay of the bias over time since K is between 0 and 1. Due to this exponential decay, the biggest impact of missing October data, relative to AK, was November 2025, and subsequent monthly estimates will have rapidly decreasing impacts.

7. Theoretical and Empirical Results

As part of the analysis of the reliability of the shift composite estimator, the theoretical results from sections 5 were compared to empirical results using historical CPS microdata and replicate weights to produce direct variance estimates for the major labor force groups of total employed and total unemployed, as shown in Table 1.

Estimate	Theoretical variance and standard error ratios		Empirical variance and standard error ratios	
	Variance ratio	Standard error ratio	Variance ratio	Standard error ratio
Total Employed	1.31	1.14	1.28	1.13
Total Unemployed	1.07	1.03	1.12	1.06

Table 1: Theoretical and Empirical Variance and Standard Error Ratios (BLS 2025)

The empirical variance and standard error ratios in Table 1 are computed for both the usual AK composite and the shift composite estimators from October 2005 to October 2025, and the results are averaged over that time period. Replicate factors account for all the complexity of the CPS sample design without relying on the assumptions in the theoretical variance derivations. Note the both the theoretical and the empirical results support the expectation of larger variances regardless of estimate type.

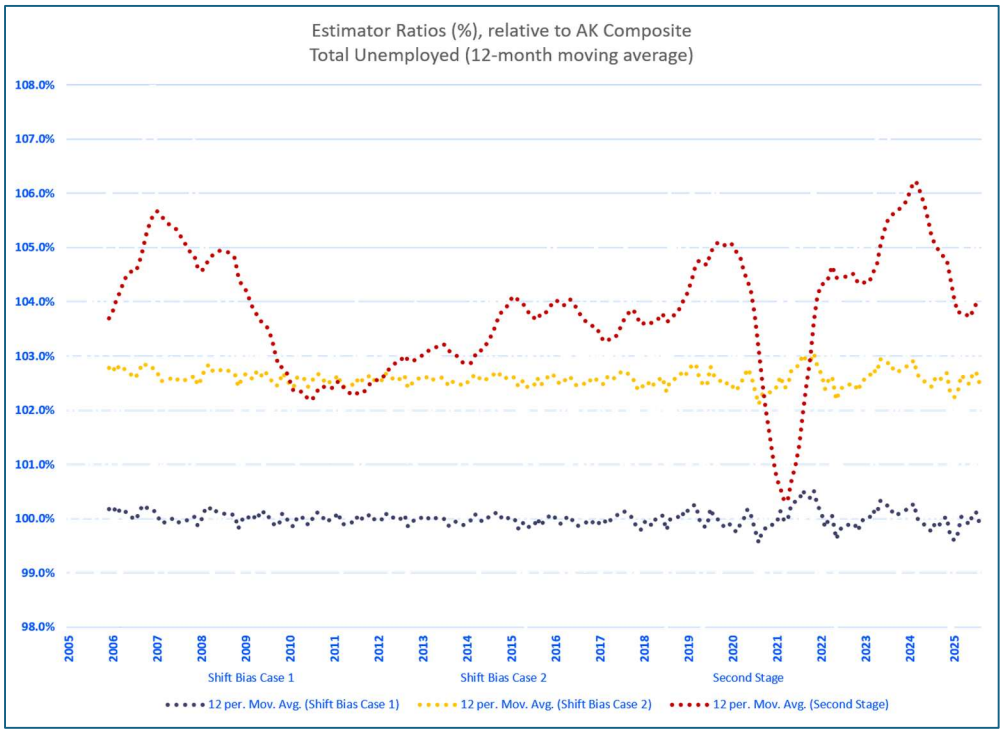


Figure 3: Estimator ratios relative to the AK Composite Total Unemployed

Figures 3 and 4 compare the two bias cases of the shift composite estimator (section 6) and the second-stage estimates, or relative to the AK composite estimator. Thus, it's important to note that since this is relative to the AK composite, the second-stage estimates, or $\hat{Y}_t^{SS}$ in Equation (1.1), denoted in red, are regularly above 100 percent as the AK composite includes a negative bias in labor force levels due to unequal compositing coefficients and differing MIS effects.

Both figures 3 and 4 illustrate that the case 2 bias assumption is consistently slightly higher than the case 1 assumption regardless of series type of employed or unemployed. Case 1, even if unlikely under the conditions of November 2025 data collection, reinforces that the shift composite estimator is unbiased relative to the AK estimator under ideal conditions. Combined, these two lines provide a range of reasonable expectations of the bias impact of the shift composite estimator, with intuitive inclination toward the higher relative bias from case 2 as being more likely.

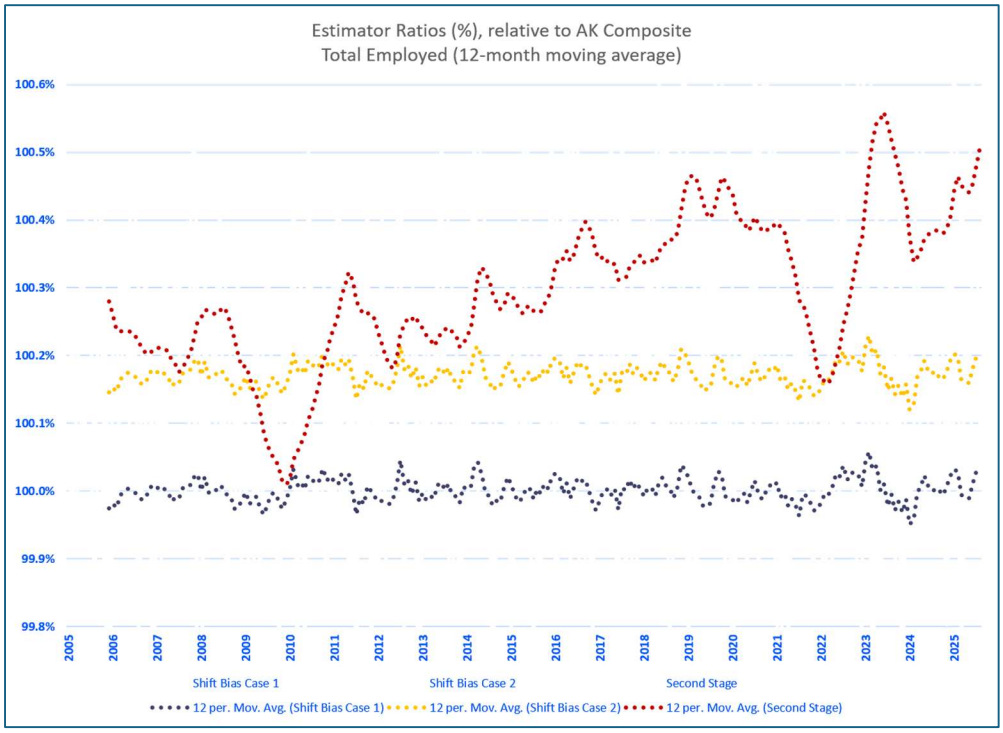


Figure 4: Estimator ratios relative to the AK Composite Total Employed

Lastly, Figure 5 shows a selection of seasonally adjusted unemployment rates for some key demographic groups. While all results above refer to not seasonally adjusted results, it is interesting to note some slight increases in unemployment rates among a few of these groups in November 2025 and slight declines after. This is simply an observational note, but it does comport with the results of this paper that a slight upward bias would be expected in unemployment estimates in the first month after the shutdown, and that such bias would quickly fade away in the following months.

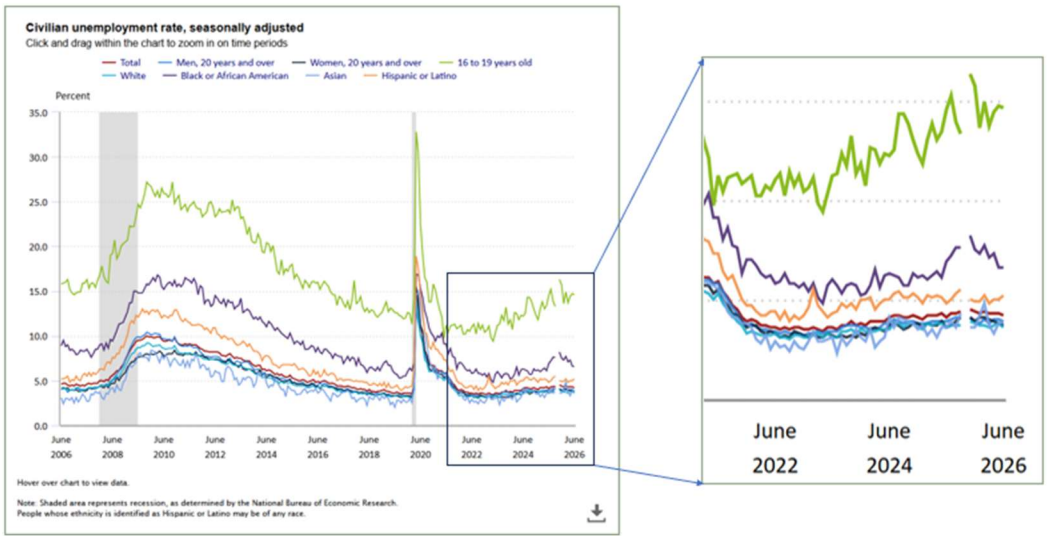


Figure 5: Civilian unemployment rate, seasonally adjusted (BLS 2026)

Theoretical and empirical results were consistent in demonstrating that the shift composite estimator was the best option to minimize adverse impacts on the time series but would incur a slight upwards variance penalty and possible bias upwards penalty. However, this is still a better alternative to restarting the usual AK composite estimator or alternatives such as the gap estimator.

The results summarized in this paper and also presented in the CPS documentation about the October 2025 government shutdown, helped inform the decision-making around difficult estimation questions in a situation unprecedented in the Current Population Survey's long history of measuring the American labor force, minimizing disruptions to the historical time series and variance inflation while maintaining methodological transparency.

References

Current Population Survey Composite Estimation, November 2025, Technical Documentation (2025). https://www.bls.gov/cps/methods/cps-composite-estimation-november-2025.pdf. U.S. Bureau of Labor Statistics.

U.S. Census Bureau (2006). Design and Methodology, Current Population Survey, Technical Paper 66 (2006). Washington, DC: authors.
—Design and Methodology, Current Population Survey, Technical Paper 77 (2019). Washington, DC: authors.

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.

Graphics for Economic New Releases, Civilian Unemployment Rate (2026). https://www.bls.gov/charts/employment-situation/civilian-unemployment-rate.htm#. U.S. Bureau of Labor Statistics.

Huang, E. and Ernst, L. (1981). "Comparison of an Alternative Estimator to the Current Composite Estimator in CPS," in Proceedings of the Section on Survey Methods Research, American Statistical Association, pp. 303 – 308.

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.

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