Comparisons of CPS Unemployment Estimates by Rotation Panel

Yang Cheng[1], Michael D. Larsen[2], Alexander F. Wakim[3]
[1] U.S. Census Bureau[1], Washington, D.C. 20233-0001
[2]George Washington University, Rockville MD 20852
[3]University of North Carolina, Chapel Hill, NC 27599-3260

Abstract: The Current Population Survey (CPS) is a monthly household survey of 72,000 households conducted by the U.S. Census Bureau for the U.S. Bureau of Labor Statistics to measure employment, unemployment, and other characteristics of the civilian non-institutionalized population in the United States. The CPS began in 1940 and provides data for key economic indicators. In this paper, we will study CPS rotation panel bias, investigate whether the estimates of unemployment for different Month-in-Sample (MIS) panels are statistically significant, and explore the assumptions underlying the AK composite estimate. We are particularly interested in the comparison of MIS 1 versus MIS 5 because they are based on personal interviews and are key elements in AK composite estimate. We apply a nonparametric statistical method with dependent permutations across time to real CPS data from 2006-2010.

Key Words: Current Population Survey, Rotation Panel Sample Design, Month-in-Sample, Rotation Bias, Wilcox Signed Rank Test, AK Composite Estimate

1. Background

The Current Population Survey (CPS) is a household sample survey, which is jointly sponsored by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (BLS). CPS is the primary source of labor force statistics (LFS) for the population of the United States. The CPS is the source of numerous high-profile economic statistics, including the monthly national unemployment rate, and provides data on a wide range of issues relating to employment and earnings. The CPS is also a source of information not only for economic and social science research, but also for the study of survey methodology. The principal purpose of CPS is to report timely and accurately on the labor force data. The U.S. Census Bureau applies a probability sample to draw about 72,000 occupied households per month for the CPS. We conduct data collection during the calendar week that includes the 19th of the month. The questions refer to activities during the prior week, that is, the week that includes the 12th of the month. The CPS also classifies adults (persons 16 years of age and older) in the civilian noninstitutional population as in the labor force (either employed or unemployed) or not in the labor force. The precise definition of employment is more detailed, but generally, the term refers to persons who (during the reference week for the survey) did one hour or more of work for pay. The precise definition of not-in-labor-force is more detailed, but generally, the term refers to persons not employed who are not looking for work. The precise definition of unemployed is more detailed, but generally, the term refers to persons not employed who want work and are available to work, and who also has to make specific recent efforts to find employment. The civilian labor force is the sum of employed and unemployed

[1] *This report is released to inform interested parties of research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.*

persons. Public debate may reference other measures of labor force underutilization, and CPS does publish several alternate measures, but the term “unemployment rate” is reserved for one specific type of ratio. The official national unemployment rate is defined as national unemployed divided by the national civilian labor force, and similar ratios can be used to compute unemployment rates by demographics or for subnational areas. Subsequent to the start of the CPS in 1940, only minor changes were made to the definitions of employed, unemployed, not-in-labor-force, and the unemployment rate.

The CPS has a complex sample survey design that allows the production of national and state data with specified reliability criteria. There is a first-stage sample of geographic units called Primary Sampling Units (PSUs). All major metropolitan areas are selected with certainty, and elsewhere a probability sample of PSUs is selected. In the second stage of sampling, a probability sample of housing units is taken within each of the selected PSUs.

The CPS is a panel rotation survey, and for any given month, 8 panels are included, each with 1/8 of the total sample. The panels are also called “rotation groups” or simply “rotations” since they are rotated in and out of active data collection. Housing units are selected from all 50 states and the District of Columbia for 4 consecutive months, out for 8 months, and then return for another 4 months before leaving the sample permanently. We call this a 4-8-4 rotation panel design. It ensures a high degree of continuity from one month to the next, one quarter to the next, as well as one year to the next. The 4-8-4 sampling scheme has the added benefit of allowing the constant replacement of the sample without excessive burden to respondents. The sample is updated each month by rotating out two panels and rotating in two panels. Each month one panel from the previous month is permanently dropped out since it finished its last month in the sample. It is replaced by introducing a new panel or 1/8 of the sample; it is rotated in for its first “month-in-sample”. Also, each month a previous month panel is temporarily dropped and is replaced by one panel that is brought back for its fifth “month-in-sample” interview. The other 6 panels just continue from the previous month and are in the second, third, fourth, sixth, seventh, and eighth months of data collection (previous month in first, second, third, fifth, sixth, and seventh month of data collection, respectively). So each month two panels are dropped or rotated out, those are replaced by adding or rotating in two panels, and ¾ of the sample (6 of 8 panels) is in common or “overlaps” with the previous month. The month-to-month sample overlap can be exploited to improve estimates of month-to-month change.

Table 1 illustrates the rotation structure. We name the rotations for clarity of exposition. Each sample is split into eight rotations (labeled here A-H). As presented here, sample 90 B has its first month of interview in January of 2012. Sample 90 B has its second, third, and fourth months of interview in February, March, and April of 2012. Sample 89 A has its fifth through eighth months in the survey from April to July of 2012. The fifth month in survey occurs one year after the first month in survey due to the eight month gap after the fourth interview month. In any given month, data are available from households that began in eight different rotations. In January of 2012, rotations 88 (C, D, E, and F), 89 (G and H), and 90 (A and B) provide data. From month to month, there are six rotations in common. The months January and February of 2012 have rotations 88 (D, E, and F), 89 H, and 90 (A and B) in common. From year to year, there are four rotations in common. Rotations 89 (G and H) and 90 (A and B), in bold, appear both in January of 2012 and January of 2013.

Table 1: Illustration of months in sample by year and month for different CPS samples and rotations. Each sample has eight rotations labeled here as (A-H).

Year and		Sample 88						Sample 89								Sample 90								
Month		C	D	E	F	G	H	A	B	C	D	E	F	G	H	A	B	C	D	E	F	G	H	A
12	Jan	8	7	6	5									**4**	**3**	**2**	**1**							
	Feb		8	7	6	5									4	3	2	1						
	Mar			8	7	6	5									4	3	2	1					
	Apr				8	7	6	5									4	3	2	1				
	May					8	7	6	5									4	3	2	1			
	Jun						8	7	6	5									4	3	2	1		
	Jul							8	7	6	5									4	3	2	1	
	Aug								8	7	6	5									4	3	2	1
	Sep									8	7	6	5									4	3	2
	Oct										8	7	6	5									4	3
	Nov											8	7	6	5									4
	Dec												8	7	6	5								
13	Jan													**8**	**7**	**6**	**5**							
	Feb														8	7	6	5						

2. Estimation

The Current Population Survey (CPS) is a multistage probability sample of housing units in the United States. It produces monthly labor force and related estimates for the total U.S. civilian non-institutional population and for various age, sex, race, and ethnic groups. In addition, estimates for a number of other population sub-domains of the nation are produced on a monthly, a quarterly, or a yearly basis. Each month a sample of eight panels is interviewed, with demographic data collected for all occupants of the sampled housing units. Labor force data are collected for persons 15 years and older. Each rotation group is a representative sample of the U.S. population. The labor force estimates are derived through a number of weighting steps in the estimation procedure. In addition, the weighting at each step is replicated in order to derive variances for the labor force estimates. In this section, we discuss the CPS estimation procedure.

Distributions of demographic characteristics derived from the CPS sample in any given month will be somehow different from the true distributions even for such basic characteristics such as age, race, sex, and Hispanic ethnicity. These particular population characteristics are closely correlated with labor force status and other characteristics estimated from the sample. Therefore, the variance of sample estimates based on these characteristics can be reduced by the use of appropriate weighting adjustments when the sample population distribution is brought as closely into agreement as possible with the known distribution of the entire population with respect to these characteristics. This is accomplished by means of ratio adjustments. There are five ratio adjustments in the CPS estimation process: the first-stage ratio adjustment, the national coverage adjustment, the state coverage adjustment, the second-stage ratio adjustment, and the composite ratio adjustment.

The main interest of CPS is the estimation of monthly totals of a given set of characteristics and their monthly changes. Let Y_t denote the unknown total for y at month t and let $\hat{Y}_{t,i}$ be the estimate of Y_t based on data in the ith rotation panel at month t, $i = 1, \ldots, 8$. As we described above, there are 7 steps of weighting adjustment to obtain $\hat{Y}_{t,i}$. These steps came from 3 categories: nonresponse adjustment, coverage adjustment (state, national), and calibration adjustment (state, race, ethnicity, age, and sex).

In the first-stage ratio adjustment, weights are adjusted so that the respective distributions of Black alone and non-Black alone census population from the sampled PSUs in a state corresponds to the Black alone and non-Black alone population distribution from the census for all PSUs in the state. In the national coverage ratio adjustment, weights are adjusted. So that the distribution of age-sex-race-ethnicity groups match independent estimates of the national population. In the state coverage ratio adjustment, weights are adjusted so that the distributions of age-sex-race groups match independent estimates of the state population. In the second-stage ratio adjustment, weights are adjusted so that aggregated CPS sample estimates match independent estimates of population in various age/sex/race and age/sex/ethnicity cells at the national level. Adjustments are also made so that the estimated state populations from CPS match independent state population estimates by age and sex.

The raking ratio estimate is

$$\hat{Y}_{t,i} = \sum_{j \in S_{ti}} b_{ti,j}\, y_{ti,j}$$

where $b_{ti,j}$ are the calibrated weights, defined to minimize the distance measure

$$\sum_{j \in S_{ti}} b_{ti,j}\, ln\left(b_{ti,j}/w_{ti,j}\right)$$

subject to $\sum_{j \in S_{ti}} b_{ti,j}\, x_{ti,j} = X_{ti}$ (the calibration equation). Here, X_{ti} is an auxiliary variable called the population control variable.

After weighting and calibration, we have the final form of estimation for CPS, which is generally called the Ratio Estimate or Second Stage Estimate in the papers related to the CPS:

$$\hat{Y}_t^1 = \frac{1}{8} \sum_{i=1}^{8} \hat{Y}_{t,i}$$

Once each record has a second-stage weight, an estimate of level for any given set of binary characteristics identifiable in the CPS can be computed by summing the second stage weights for all the sample cases that have that set of characteristics. This type of estimate has been variously referred to as a Horvitz-Thompson estimate, a two-stage Ratio Estimate, or a simple weighted estimate. However, the estimate actually used for the derivation of most official CPS labor force estimates that are based upon information collected every month from the full sample is a composite estimate.

Each panel or rotation by itself should provide estimate, $\hat{Y}_{t,i}$, of the true size of the labor force and the number of persons employed and unemployed. Each month responses are gathered from eight panels that are now in month-in-sample 1, 2, 3, 4, 5, 6, 7, and 8. There are many steps in estimation, but each panel estimate in the current month is more-or-less an average of 8 panel estimates, $\hat{Y}_t^1$. A full review of CPS weighting and estimation is beyond the scope of this paper. Interested readers can find details in Chapter 10 of Technical Paper 66.

The situation is that every month each panel/rotation should ideally estimate the same quantities without differences in bias. It is natural that sample estimates have sampling variability, but bias is the issue. Previous studies uncovered systematic relative biases between panels, and as will be seen this paper offers further evidence. However, the purpose of this article is not to formally re-investigate all of the systematic differences, but to discuss how one would assess if parameters describing the sampling distribution of estimators are different by month-in-sample. For example, focusing on unemployment as an example and the mean as a summary, what evidence is there that estimators of the number of unemployed in a given month have different expected values by rotation?

3. Evidence for differences by month-in-sample (MIS)

We ask ourselves why makes each panel estimate, $\hat{Y}_{t,i}$, looks differently. We can list following reasons:

- Rotation panels in different months may be drawn from different frames
- Survey interviewers collect data in survey months one and five in-person using Computer Assisted Personal Interviewing (CAPI) whereas re Assisted Telephone Interviewing (CATI) is used for most interviews in the other six survey months.
- Survey months one and five also initiate data collection after a period of no data collection. In the months other than one and five, survey respondents can reply that their status has or has not changed from the previous month.
- Some respondents to early survey months might refuse or not be available for interview in later months. Other respondents move.
- Another reason for months in survey 1 and 5 having higher expected values is that new growth is added only at those months.

Before we conduct statistical test among panel estimates, we narrow our scope to unemployment variable only. Also, let us focusing on comparison in survey month one versus five by applying CPS data set from January 2006 to December 2010. There are a total 60 months CPS data.

Figure 1 displays estimated monthly numbers of unemployed, $\hat{Y}_{t,i}$, for the eight panels/rotations from January 2006 to December 2010, as well as average of eight panel estimates, $\hat{Y}_t^1$. In general, all panel estimate, $\hat{Y}_{t,i}$, and average of 8 panel estimates, $\hat{Y}_t^1$, look like having the same pattern. Unemployment was rising nationwide during this time period. The 8 panels for a month are identified as MIS 1 through MIS 8 (month-in-sample 1 through 8). MIS 1 estimate is above other MIS estimate and average of MIS estimates. Except MIS 1 estimate, everyone else and the rest are all bunched together. MIS5 estimate appears to be in about the middle – half are above and half are below. Although not the official CPS estimation method, the mean of the eight monthly panel estimates is one way to estimate the number of unemployed in a given month. Based on this visual display, if we try to look extremely hard, we may see that rotation MIS 1 estimate and maybe perhaps rotation MIS 2 estimate seem to be generally above the average, whereas estimates from rotations MIS 8, MIS 7, and to a lesser degree MIS 6 seem to be generally below the average.

Figure 1: Comparison among 8 panel estimates and their average for national unemployment

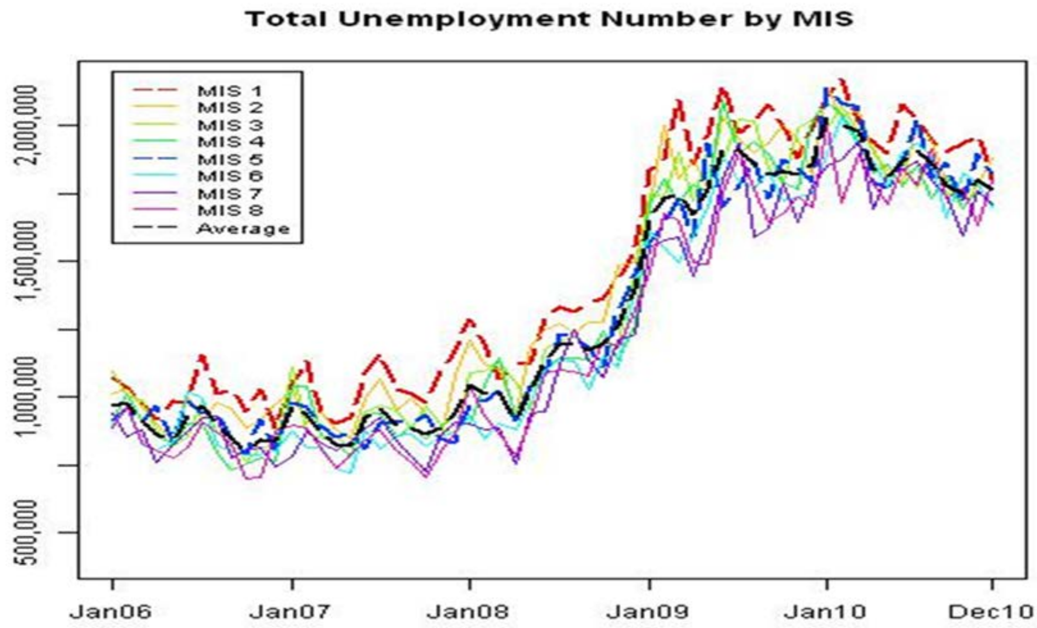
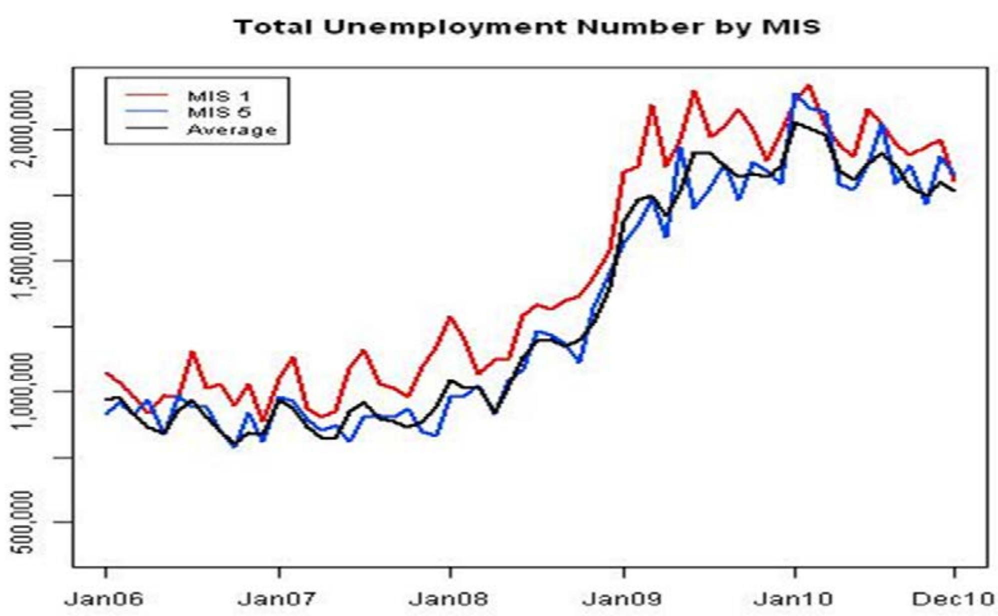


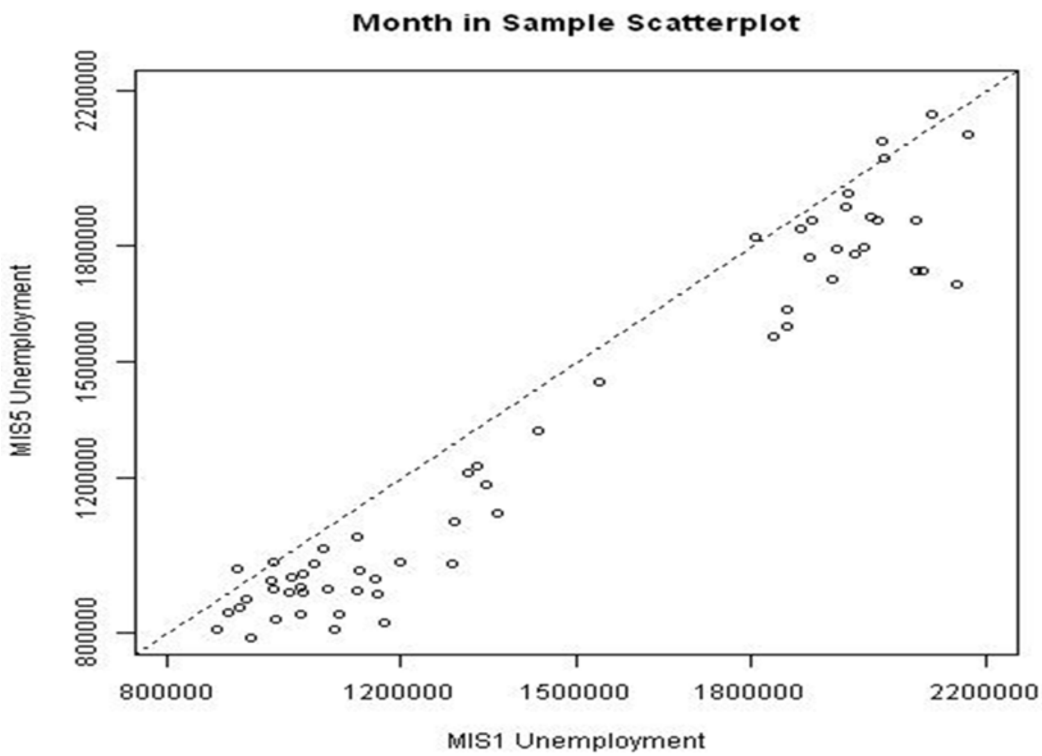
Figure 2: Estimated monthly numbers of unemployed for Rotations 1 and 5 from January 2006 to December



Of particular interest to CPS are rotations in month-in sample 1 and 5 (MIS 1 and MIS 5) because the survey environment is different from for the other rotations. For example, rotations 1 and 5 are phase-in rotations and both have a large proportion of interviews conducted in person. **Figure 2** more clearly shows that MIS 5 estimates is quite similar to average panel estimates and that MIS 1 estimate is above MIS 5 estimate and average panel estimates. From this graph, it is pretty clear that there is not a MIS5 bias comparing to average. However, is the MIS 1 bias suggested by this graphs significant?

Figure 3 is another graphical representation to displays estimated numbers of unemployed for these two rotations. This graphs shows MIS 1 estimate along the x-axis where MIS 5 estimate along the y-axis. Most of the points are below the line $y = x$, which suggests that MIS 1 estimate is consistently higher than MIS 5 estimate for the same month. As we count carefully, rotation MIS 1 estimate is higher than MIS 5 estimate in 55 out of 60 months.

Figure 3: Unemployment estimate for rotations 1 is higher in 55 out of 60 months than unemployment estimate for rotations 5 from January 2006 to December



4. 5 methods of comparison

4.1 Simple Binomial

A null hypothesis that these two months should have equal estimates translates into a probably of ½ that rotation MIS 1 estimate is higher than rotation MIS 5 estimate in a specific month, that is, null hypothesis H_0: $\Pr(Y_1 \geq Y_5) = 0.5$. Now, we can treat this hypothesis test problem as the same as tossing a fair coin: a binomial random variable based on 60 independent trials with a probability of ½ of each of two possible outcomes would produce 55 or more results of one kind is extremely small; a two-sided p-value for such a binomial test would be less than 10^{-10}. This odd is smaller than the probability for wining a lottery. Thus, we reject null hypothesis.

4.2 Simple Binomial with use of independent months

The calculation just described, however, is not strictly appropriate or conclusive, because the 60 months do not have independent estimates. Problem is here that we assume 60 independent trials. Nevertheless, there is a lot of time series structure. Over the course of 60 months, forty-eight rotation panels appear in survey month one and then twelve months later in survey month five: data gathered at two time points on the same subjects are dependent. One way around this is to base the comparison on only twelve consecutive months. A single year has independent samples in rotations 1 and 5, and there is no overlap. To produce a more powerful test, one could combine data from months 1 through 12, 25 through 36, and 49 through 60. Skipping a year avoids overlap. Similarly, we can combine data from months 13 through 24 and 37 through 48 and not violate independence. The exact p-values concerning the null hypothesis of p=1/2 can be computed using exact binomial probabilities in **section 4.1**. Results are shown in **Table 2**. Except for the last twelve months taken separately, the results are overwhelmingly significant and all in one direction; rotation 1 has higher unemployment estimate than rotation 5. A 95% exact confidence interval for the proportion of times that rotation 1 is higher than rotation 5 based on months 1-12, 25-36, and 49-60 is (0.71, 0.95). A 95% exact confidence interval based on months 13-24 and 37-48 is (0.86, 1.00). Corresponding to the small p-values, both intervals exclude 0.50.

Although evidence based on the graph and the binomial test described above seem quite clear, examination of the graph and the binomial test both omit some information (e.g., the exact values of the quantitative estimates) that could be used to compare rotations. What else can be done to assess where the difference between rotation 1 and rotation 5 is statistically significant?

Table 2: Number of times the unemployment estimate in rotation 1 is larger than in rotation 5 for sixty months divided into five year-long blocks.

Months	Number of Months	# Times Rotation 1 > 5	P-value (two sided)		Months	Number of months	# Times Rotation 1>5	P-value (two sided)
1-12	12	10	0.0386		1-12, 25-36, 49-60	36	31	<0.0001
13-24	12	12	0.0005					
24-36	12	12	0.0005					
37-48	12	12	0.0005		13-24, 37-48	24	24	<0.0001
49-60	12	9	0.1460					

4.3 Quantitative comparison using Normal or T test

The estimates of unemployment are quantitative random variables. The average difference between the estimates in rotations 1 and 5 is a linear contrast in these random variables. The formula below divided by 60 gives the value of the average difference. In the example comparing rotations one and five, the value is 141,565.

$$\sum_{t=1}^{60}\hat{Y}_{t,1}-\sum_{t=1}^{60}\hat{Y}_{t,5}=\sum_{t=1}^{48}(\hat{Y}_{t,1}-\hat{Y}_{t+12,\,5})+\sum_{t=49}^{60}\hat{Y}_{t,1}-\sum_{t=1}^{12}\hat{Y}_{t,5}$$

The variance of a linear contrast in random variables involves the variances of each random variable and the respective covariance. If survey participants are respondents in both rotation 1 and rotation 5 (twelve months later), then the estimates are correlated and the covariance are not zero. If estimates in panels 1 and 5 are independent for a given month *t*, then covariance are zero. The variance then is given in the formula below, where Vstands for variance, and the standard error of the average difference is computed by dividing the variance by 60 and then taking the square root.

$$V\left(\sum_{t=1}^{60}\hat{y}_{1t}-\sum_{t=1}^{60}\hat{y}_{5t}\right)=\sum_{t=1}^{48}V\left(\hat{y}_{1t}-\hat{y}_{5,t+12}\right)+\sum_{t=49}^{60}V(\hat{y}_{1t})+\sum_{t=1}^{12}V(\hat{y}_{5t})$$

The variance of the difference $\hat{y}_{1t}-\hat{y}_{5,t+12}$can be computed as

$$V\left(\hat{y}_{1t}-\hat{y}_{5,t+12}\right)=V(\hat{y}_{1t})+V\left(\hat{y}_{5,t+12}\right)-2C\left(\hat{y}_{1t},\hat{y}_{5,t+12}\right),$$

where *C* is the covariance.

Figure 4: Histogram of differences (in 100,000s of people) between estimates in rotation 1 and 5 in sixty months.

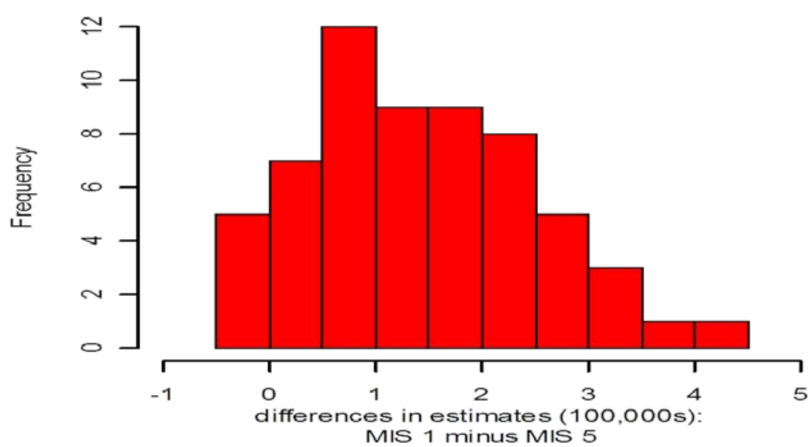


Figure 4 shows a histogram that the individual monthly differences are roughly normally distributed, but slightly skewed to the right. Since differences are roughly normally distributed, we apply either Normal or T test for difference between the estimates in rotations 1 and 5.

The standard error of the linear combination statistic is 69,000. By the Central Limit Theorem (CLT), the estimate divided by its standard error should be distributed approximately as a normal random variable. The standardized value is 2.05 = 141,565/69,000. Given the distribution of the individual differences, the CLT argument for judging the value 2.05 for the linear combination probably is reasonable. Comparing to a normal distribution, a two-sided p-value for testing the null hypothesis that there is no difference has a value of 0.040. If a t-distribution with 59 degrees of freedom (59=60-1) is used, the p-value is 0.044. Both are less than the usual 0.05 level cut off, but not by much.

4.4 Wilcoxon Signed-Rank test

Nonparametric statistical methods evaluate ranks instead of original quantitative measurements to avoid distributional assumptions associated with parametric methods. The Wilcoxon Signed-Rank test for comparing data from two matched groups is relevant here. In this procedure, one takes the difference (MIS 1 minus MIS 5) in quantitative values in a given month. This produces sixty month-specific differences if applied to the full set of estimates. The null hypothesis is that the center of the distribution of differences is zero. The statistic is computed by ranking the size of the absolute differences and then summing the ranks when the MIS 1 estimate is larger than the MIS 5 estimate. In this application, there are no ties. The statistical significance of the observed sum of ranks is assessed by considering permutations of the data under the null hypothesis. If there are n pairs of observations, then there are 2^n possible permutations. For each month, if there is no difference in distribution for MIS 1 and MIS 5, then whether MIS 1 or MIS 5 is the larger of two values recorded in a month should be determined by pure chance. If the observed value of the statistic is extreme in comparison to the distribution determined by permuting value of MIS 1 and MI 5, then the null hypothesis is not likely to be true.

Table 4: Illustration of the Wilcoxon Signed-Rank Test Statistic for actual data in five months (top) and with one permutation under the null hypothesis (bottom).

Month	MIS 1	MIS 5	Difference	Absolute Difference	Rank of Abs. Diff.	Is MIS 1 > MIS 5?	Contribution to sum
1	1,074,210	913,450	160,750	160,750	5	Yes	5
2	1,031,280	954,290	76,990	76,990	3	Yes	3
3	981,050	913,490	67,560	67,560	2	Yes	2
4	918,210	965,490	-47,280	47,280	1	No	0
5	985,460	837,390	148,060	148,060	4	Yes	4
Rank Sum for actual data							14

Month	MIS 1	MIS 5	Difference	Absolute Difference	Rank of Abs. Diff.	Is MIS 1 > MIS 5?	Contribution to sum
1-permuted	*913,450*	*1,074,210*	-160,750	160,750	5	*No*	0
2	1,031,280	954,290	76,990	76,990	3	Yes	3
3	981,050	913,490	67,560	67,560	2	Yes	2
4	918,210	965,490	-47,280	47,280	1	No	0
5	985,460	837,390	148,060	148,060	4	Yes	4
Rank Sum for permuted data							9

The computation of the statistic is illustrated in **Table 4** using 5 months of data. First, the statistic is computed for five months of actual data. The sum of the signed ranks is 14. Second, the statistic is computed for one permutation of the data, switching MIS 1 and MIS 5 values in the first month. The rank of the absolute difference is the same in month 1 under the original data and under the permuted row. The rank of the absolute difference (5) does not contribute to the sum under the permuted data, because after permutation MIS 1 is less than MIS 5. As a result, the sum of signed ranks is 9 for the permutation example. Thus, the actual data have a larger rank sum than the permuted data in this case. In all $2^5=32$ permutations, the actual data achieve the 2nd highest rank sum. Two out of 32 is 0.0625 (a 1-sided p-value) and 4 out of 32 (a 2-sided p-value) is 0.1250. In this illustration with only 5 months, the result would not be judged statistically significant, but there are many more months of data available.

When analyzing the original sixty months of data, the problem of non-independence arises here as it did with the first method considered. There are 48 rotations that contribute to both a first month in sample and a fifth month in sample; there are not sixty independent differences for use in the test. To get around the problem, one can use subsets of the data such as those used in **Table 2** that are independent. **Table 5** presents results for the Wilcoxon Signed-Rank Test on subsets of the data. All differences are statistically significant at the 0.05 level. The result from the last 12-month period might not be considered strong evidence because we have conducted five tests for 12-month periods, but it still suggests that rotation 1 reports higher unemployment than rotation 5.

There are additional theoretical problems with using the Wilcoxon Signed-Rank Test. One assumption is that MIS 1 and MIS 5 estimates are uncorrelated for the same month, whereas some correlations are induced by CPS sampling and weighting. Another assumption is that the sampled observations have the same standard error, but 1) small differences in MIS 1 and MIS 5 variances would arise from the systematic differences between the two MIS and 2) variance generally slightly increases over the period of rising unemployment used for this study.

Table 5: Wilcoxon Signed-Rank Test Statistic (WSRT) for comparing the unemployment estimate in rotation 1 to the estimate in rotation 5 for sixty months divided into five year-long blocks.

Months	Number of Months	Statistic WSRT	P-value (two sided)		Months	Number of months	Statistic WSRT	P-value (two sided)
1-12	12	75	0.0024		1-12, 25-36, 49-60	36	643	<0.0001
13-24	12	78	0.0005					
24-36	12	78	0.0005					
37-48	12	78	0.0005		13-24, 37-48	24	300	<0.0001
49-60	12	68	0.0210					

4.5 A Nonparametric Procedure Applied to All Sixty Months (Permutations)

It is reassuring that the graphical plus four numerical methods considered so far give consistent and unequivocal results. Still, it seems unfortunate that the two nonparametric methods have discarded some data, or cannot be applied to data from all sixty months together. Is there an alternative that would use all of the data but not rely on normality via the CLT for evaluating distributions of test statistics? Several departures from the theoretical ideal have been mentioned, however a major problem with using all the 60 months together is that rotations in MIS 1 appear again twelve months later in MIS 5 – a problem of non-independence.

A sharp null hypothesis by definition is one that determines the sampling distribution of a statistic when the null hypothesis is true. The sampling distribution is sometimes called a reference distribution for judging how extreme a statistic value is. What does a null hypothesis that MIS 1 and MIS 5 have exactly the same distribution in a given month imply when considering the 4-8-4 rotation pattern? First, under the null hypothesis there should be no expected difference between the MIS 1 and MIS 5 estimates in a given month; one could switch them and on average, it should not matter. Second, because each rotation contributes once to the MIS 1 estimate and once to the MIS 5 estimate, a flip of MIS 1 and MIS 5 at one month has a cascading effect. That is, if one permutes the MIS 1 estimate to the MIS 5 column, then one should also permute the MIS 5 column for this rotation twelve months later to the MIS 1 column. It would seem only fair when evaluating a statistic that a rotation that generally had higher employment on average should appear once in MIS 1 and once in MIS 5.

Figure 5: Illustration of Permutations in 28 months under the Sharp Null Hypothesis with 4-8-4 constraints

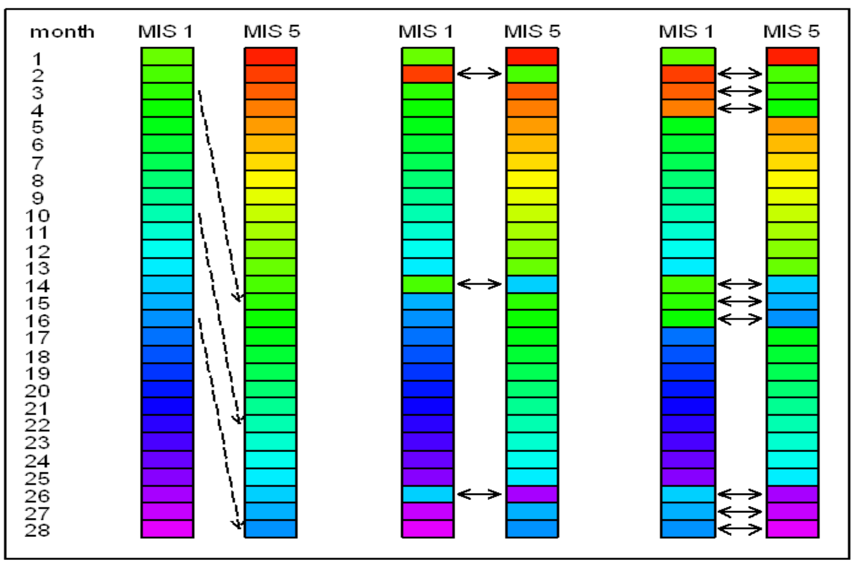


Figure 5 illustrates the permutation-with-dependence procedure on 28 months. Each month has a MIS 1 sample and a MIS 5 sample, which is linked 12 months apart (dashed arrows). The first pair of columns illustrates the original rotations. The second pair illustrates a permutation in month 2 that causes permutations 12 and 24 months later. The third pair illustrates a permutation in months 2-4 and subsequent dependent permutations.

Due to the dependence structure, there are not 2^{60} but only 2^{12} =4,096 possible permutations. Both non-parametric statistics (the Binomial test of **Table 2** and the Wilcoxon Signed-Rank test of **Table 5**) can be evaluated under the dependent randomization permutation distribution. Fifty-five out of sixty months have MIS 1 higher than MIS 5. In the 4,096 permutations, 55 is the largest value, which corresponds to a (two-sided) p-value less than 0.0005. It is not possible to put all of the largest values month-by-month into MIS 1 and produce a value of 60 due to the dependence. In addition, the largest value of the Wilcoxon Signed-Rank test is 1,799. Consequently, it is clear that there is a statistically significant difference and that MIS 1 produces larger estimates of unemployment than MIS 5.

5. Conclusion

The project presented here on examining panel bias used monthly Current Population Survey panel/rotation data, also referred to as month-in-sample (MIS) data. It is a high-profile monthly household survey producing national and subnational labor force data including the national unemployment rate, a key economic indicator. The CPS has a complex sample design and complex estimation/weighting, and that offers challenges when analyzing month-in-sample bias.

For any given month, the CPS has eight panels. One panel is surveyed the first time (MIS 1), another panel is surveyed the second time (MIS 2), and the remaining six panels are surveyed their respective third through eighth times (MIS 3 through MIS 8). Completely eliminating bias is impractical, but it is generally desirable for the panels of a survey like CPS to have the same relative bias. It is important to measure differences in panel bias or month-in-sample bias since small differences can have appreciable impacts on critical estimates. The study of CPS month-in-sample bias spans several decades and a comprehensive re-investigation was not attempted in this project, rather new ways at approaching the panel bias problem were tested.

Three innovative applications of parametric and nonparametric tests were presented: linear contrasts, Wilcoxon Signed-Rank Test, and Permutation methods. Using CPS unemployment for demonstration purposes, all five statistical tests plus a graphical examination indicate that MIS 1 unemployment is higher than MIS 5 unemployment.

The methods offer promise. Departures from the theoretical ideal were noted for the parametric and nonparametric data. Some departures to the theoretical ideal are caused, for example, due to correlations induced by the complex CPS sampling and weighting. Other departures are caused by data changes over time (e.g., rising unemployment) and the accompanying changes in variances. Substantial work is needed to fully investigate the impact of these departures on the validity and usefulness of the methods.

References

Bailar, B.A. (1975). The Effects of Rotation Group Bias on Estimates from Panel Surveys, Journal of the American Statistical Association, 70, 23-30.

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

Cheng, Y. (2012). Overview of Current Population Survey Methodology, Proceedings of the Survey Research Methods Section, American Statistical Association, 3965-3979.

Evinger, S. N. (1996). Results of CPS research on the measurement of race and ethnicity, Chance, 9 (No. 2), 53-56.

Hansen, M.H., Hurwitz, W.H., Madow, W.G., (1953). Sample Survey Methods and Theory, Vol. II, New York: John Wiley and Sons.

Hansen, M.H., Hurwitz, W.N., Nisselson H., Steinberg, J., (1955). The Redesign of the Census Current Population Survey. Journal of the American Statistical Association, 50, 701-719.

Hollander, M., and Wolfe, D.A. (1999). Nonparametric Statistical Methods, 2nd Edition Wiley-Interscience.

Larsen, M, Wakim, A, Rottach, R., and Cheng, Y (2014), Nonparametric Comparison of Current Population Survey Panel Estimates (accept by Chance).

Norwood, J.L. (1989). The employment situation: How do we know about it? Chance, 2 (No. 3), 26-29.

Plewes, T. J. (1994). Federal agencies introduce redesigned Current Population Survey, Chance, 7 (No. 1), 35-41.

U.S. Census Bureau, (2006). Design and Methodology: Current Population Survey, Technical Paper 66, October 2006, http://www.census.gov/prod/2006pubs/tp-66.pdf.