A COMPARISON OF NONPARAMETRIC METHODS WITH PARAMETRIC METHODS FOR THE CPS CATI/CAPI MODE EFFECTS ANALYSIS

Jenny Thompson and Randall Parmer
Randall Parmer, Bureau of the Census, Washington, DC 20233

KEY WORDS: Wilcoxon Test, Paired Sign Test, Quade Test

I. Introduction

The official monthly civilian labor force estimates from January 1994 onward are based on data from a comprehensively redesigned Current Population Survey (CPS). The redesign included implementation of a new, fully computerized questionnaire and an increase in centralized computer-assisted telephone interviewing (CATI). To gauge the effect of the CPS redesign on published estimates, the Parallel Survey (PS) was conducted using the new questionnaire and data collection procedures from July 1992 through December 1993. Annual average estimates from the PS were used to examine the effect of the CPS redesign on major labor force estimates.

A secondary consideration was an investigation into the possible effect of selected factors associated with the new questionnaire or collection mode on major labor force estimates. Special studies were embedded in the CPS and the PS during the same time period to provide data for testing hypotheses about the effects of these new methodological differences on labor force estimates. October 1992 through December 1993 data from these studies were used for this mode effects analysis. The results of these parametric tests are provided in reference [5].

The published mode effects analysis consisted primarily of two-sample t-tests. This test is popular because it is easily interpretable and fairly robust to the assumption of normality. The latter assumption is, however, difficult to verify with complex survey data.

Nonparametric applications for the mode effects analysis were an appropriate compliment to the parametric analysis. The purpose of the special mode effects analysis studies was to examine **contrasts** in estimates between split panels. The statistics of interest were the estimated **differences** in split panel estimates, rather than the point estimates for each panel. In fact, the intrinsic value of the panel estimates was debatable, given that the analysis used sub-national statistics. Because no meaning per se was attributed to the value of the panel estimates, binomial type and rank-based analysis were logical extensions. From a mathematical perspective, analysis lost little by using distribution-free techniques. When the normality assumption is in doubt, nonparametric tests are often more powerful than their parametric counterparts.

We applied four nonparametric tests to split panel data: the Mann-Whitney Wilcoxon, the Paired Sign Test, the Wilcoxon Signed Rank Test, and the Quade Test. Comparisons of the Mann-Whitney Wilcoxon results to the published normal theory results are provided. All tests were performed on monthly data, from October 1992 through December 1993. March 1993 data was excluded from the analysis because one of the CATI facilities was shut down during interview week due to a blizzard. In addition to testing the monthly data, we tested fourteen month averages (see [3]).

II. Hypotheses

A. Description of Split Panel Data

In both the CPS and the PS, the Census Bureau designated selected Primary Sampling Units (PSUs) as "CATI-eligible," where a PSU is a county or group of counties. Sample within these PSUs was randomly split into two representative panels: a CATI-eligible panel, and a non-CATI panel. Households in the CATI panel were eligible for centralized computer-assisted telephone interviewing (CATI) after the initial personal visit interviews, provided that the respondents had a telephone, spoke English or Spanish, and agreed to telephone interviews in subsequent months. Consequently, not all households in the CATI panel were interviewed from a centralized telephone facility. All households in the non-CATI panel were designated as ineligible for CATI interviewing.

The set of CATI-Eligible PSUs differed by survey. In addition, the hypotheses tested by each split panel differed. The CPS split panel data was used to test for a combined centralized and computer-assisted telephone interviewing effect. CPS CATI interviews were conducted with a fully computerized version of the old paper questionnaire, which had a slightly modified wording of the lead-in to the labor force question. It was therefore impossible to distinguish whether a difference in unemployment rate between split panels was due to centralization,

computer-assisted interviewing, or the slightly modified questionnaire. Parametric results from the CPS study are provided in [5] and [6]. The PS split panel data was used to test for a centralized telephone interviewing effect. All of the PS data were collected using computer-assisted interviewing with the redesigned CPS questionnaire. Parametric results from the PS study are provided in [5]. Unfortunately, the split panel design for the PS did not permit a nonparametric analysis: only one tenth of the PS CATI eligible areas was designated for the non-CATI panel.

The split panel data from the intersection of the CPS and the PS CATI-eligible areas was used to test for a third effect: the effect of the new questionnaire, given centralized telephone interviewing. In this case, estimates from the PS CATI panel were compared to estimates for the CPS CATI panel in the common areas. The "treatment" examined was the questionnaire: the PS data used the fully automated redesigned questionnaire; the CPS data used the old paper questionnaire, which was automated for CATI. Parametric results for the common PSU tests are provided in [6].

Further details of test hypotheses and split panel design and limitations are provided in [6].

B. Application of Nonparametric Tests to Split Panel Data

1. Estimates

We calculated two estimates for each PSU for each hypothesis: one estimate for the "treatment" panel, the other for the control panel. For the CPS data, the treatment panel was the CATI panel; the control panel was the non-CATI panel. For the Common CATI-Eligible PSU data, the treatment panel was the PS CATI panel; the control panel was the CPS CATI panel. PSU/panel estimates are "unbiased," i.e. baseweighted, with a weighting control factor (to adjust for subsampling in the field), and an adjustment for probability of being in the particular panel. Because first and fifth month CPS and PS interviews were never conducted from a CATI facility, the data from these months of interview were excluded from the panel estimates for testing these hypotheses.

We verified the unweighted PSU sample sizes in the split panels using a fourteen month average of data. The sample size consideration forced us to exclude the PS split panel data from our analysis: several PS CATI Eligible PSUs had non-CATI panel estimates based on one or two observations. We decided that the other two sets of data had adequate PSU/panel sample sizes to pursue this analysis. The effect of a few small PSU/panel estimates did come into play when testing fourteen month averages as described in section II.B.3.a.

Generally, the number of CPS CATI Eligible PSUs was adequate for the analysis: seventy-five PSUs are included in the fourteen month average. This number ranged from a minimum of sixty-four PSUs to a maximum of seventy-two in a given month. Moreover, the sample sizes in the two panels in the CPS CATI Eligible PSUs are fairly equitable. In contrast, the sample of Common CATI Eligible PSUs was "borderline" adequate for the analysis: fifty-two PSUs are included in the fourteen month average. The number of Common CATI PSUs each month ranged from a minimum of forty-two to a maximum of fifty-one. In addition, the sample size in the CPS CATI panel was approximately four times larger than the PS CATI panel sample in any given PSU, and so the two panel's estimates did not have comparable reliability.

Using PSU/panel estimates of levels would have weighted the analysis too heavily towards observations from the larger PSUs. Instead, we considered three different rates: Unemployment Rate, Employment to Population Ratio, and Civilian Labor Force (CLF) Participation Rate. Descriptions of these rates are provided in [7]. These three rates are the major labor force characteristics estimated monthly by the CPS.

2. Assumption Validation

To determine the alternative for the Mann-Whitney tests, we plotted the empirical CDF of both panels for all three statistics within the hypothesis data set. If they had the same shape, or roughly the same shape, then we used a location shift alternative. For example, a location shift alternative is appropriate for testing the difference in CDF by panel of unemployment rate in CPS CATI PSUs, as demonstrated by Figure 1 below. If the two CDFs did not appear to have the same shape, we tested for differences in CDF with no assumptions about shape.

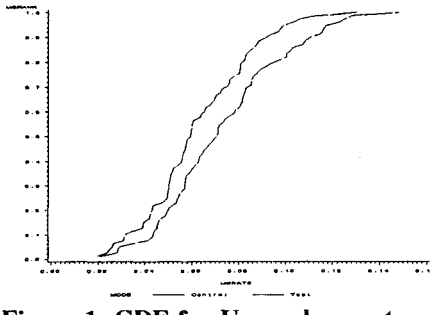

Figure 1: CDF for Unemployment Rate - CPS CATI Eligible PSUs

The Paired Sign Test assumption of independence was easily met, since the PSUs are by definition mutually independent.

We used stem-and-leaf plots of the paired PSU differences to verify the symmetry assumption for the Wilcoxon Signed Rank test. The assumption of mutual independence holds for the same reason as the Paired Sign Test. We assumed that the split panel differences within PSU have the same median, since the panels are each a random sample from the same parent sample. These data do not meet the optional assumption of constituting a random sample: all PSUs were **non-randomly** chosen for CATI eligibility, to meet specific workload criteria.

3. Fourteen Month Averages

Each fourteen month average rate is actually the ratio of two averaged estimated levels. For example, the fourteen-month-average unemployment rate used is the ratio of the fourteen-month-average estimated unemployment level divided by the fourteen month average estimated Civilian Labor Force (CLF) level. In other words, our statistics are <u>not</u> the average of the fourteen individual rates for a PSU/panel.

Each PSU/panel averaged estimate is defined as the sum of the weighted PSU/panel estimated level for each PSU's panel divided by the total months that the PSU was included in our study. The denominator could therefore be any value ranging from one to fourteen, although it was generally fourteen.

Analysis of fourteen month averages must be taken in conjunction with the monthly results. The paired data techniques are in particular sensitive to sample size. The fourteen month averages include <u>all</u> of the PSU estimates. As an extreme example, a PSU that was only in sample for a month would be included in the test statistic with exactly the same weight as a fourteen month average estimate. A small PSU with an "unusually" high difference would probably have a large rank. If the total number of PSUs is small, the average may yield a "significant" result, even though the monthly results yield consistently **non-significant** results.

4. Two-Sample Tests and Paired Data Tests

Split panel data can be examined in two ways: as two independent samples, or as a sample of paired differences. The two-sample analysis compares the difference in expected value between two distributions. This interpretation is particularly convenient for a parametric analysis of complex survey data, since it requires only two estimates of variance: one per panel.

There are analytical disadvantages of pooling the data within each panel, however. Each PSU in a complex survey design represents a particular stratum, and the set of PSUs under consideration are not homogeneous. In addition, pooling the observations in a panel could conceal a true effect. Consider this hypothetical data set:

PSU	Test X's	Test N's	Test Rate	Control X's	Control N's	Control Rate	Paired Difference (Test-Control)
1	40	80	0.500	20	80	0.250	0.250
2	60	160	0.375	40	160	0.250	0.125
3	80	240	0.333	60	240	0.250	0.083
4	9	90	0.100	0	90	0.000	0.100
5	11.5	230	0.050	80.5	230	0.350	-0.300

Because both panels have the same sample mean ($\approx .25$), the test statistic for the two-sample t-test is zero, and one would conclude that the there was no effect present. However, a consideration of the paired differences might provide some evidence to the contrary, since the mean of the paired differences (≈ 0.052) is greater than zero.

5. One-Sided Tests

Because we had prior knowledge of the direction of the expected differences, we used one-sided tests for the Paired Sign Test and for the Wilcoxon Signed Rank test. The CATI Phase-In Study described in [6] had repeatedly shown a positive effect on the unemployment rate, i.e. including CATI interviewing yielded a higher unemployment rate. As described in [7] and [8], the new questionnaire had been designed to improve major labor force estimates. We therefore expected larger unemployment rates, employment to population ratios, and CLF Participation Rates for the treatment panels.

III. <u>Results</u>

Results are discussed by hypothesis.

A. Tests for a Combined Centralized and Computer-Assisted Interviewing Effect CPS CATI Phase-in Project Data

1. Unemployment Rate

Table One summarizes the nonparametric test results for unemployment rates using fourteen month averages.

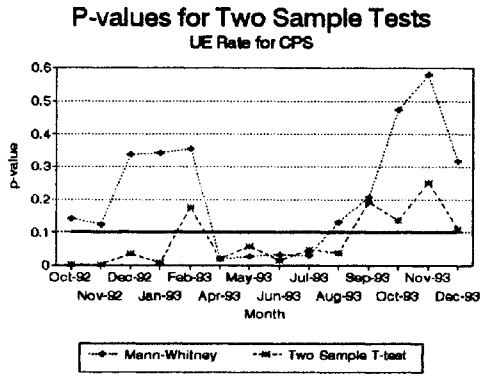

Figure 2

As seen in Figure 1 (II.B.2.), the location shift alternative is appropriate for the Mann-Whitney test. This test reinforces the two-sample t-test results. In fact, the PSU/panel unemployment rates tested as normally distributed, and so the two-sample results are consistent: the Mann-Whitney rejects the hypothesis of no difference in distribution function, but the significance level is not nearly as high. Note the consistency between the t-test and Mann-Whitney test results in the monthly p-value plots provided in Figure 2.

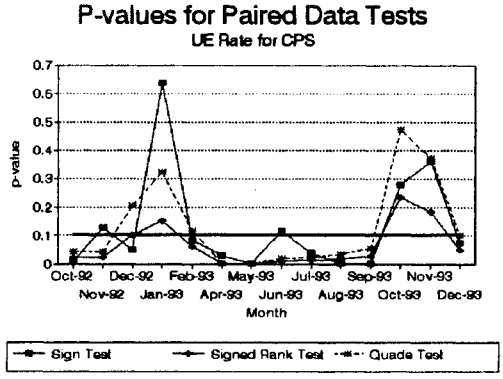

Figure 3

The results of the paired data tests provide more evidence for this CATI effect on unemployment rate. First, the paired sign test, generally not a very powerful test, has a highly significant p-value even for a one-sided test. The other, more powerful paired data tests have even smaller p-values. Finally, the p-value plots for the paired data unemployment rate tests reinforce the results. Figure 3 contains the p-value plots for these tests.

2. Employment to Population Ratio

None of the tests provided any evidence of a CATI effect for employment to population ratio.

3. CLF Participation Rate

Table Two summarizes the nonparametric test results for CLF participation rates using fourteen month averages. A priori, we expected a **positive** effect on the CLF participation rate. There is strong evidence from this study and from the study presented in [6] of a positive effect on the unemployment rate, and no evidence of an effect on the employment to population ratio. Because the CLF participation rate is a linear combination of these two statistics, we expected an overall positive effect when this type of CATI interviewing was included.

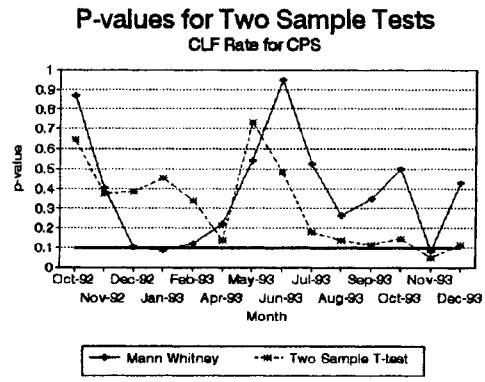

Figure 4

Both two-sample tests using the fourteen month average present very consistent results. Both would have p-values slightly smaller than 0.10 for a **one-sided** test, assuming that CLF participation rate increased when CATI interviewing was included, thus showing very preliminary evidence of such an effect. Our CDF plots reinforced the Mann-Whitney conclusions.

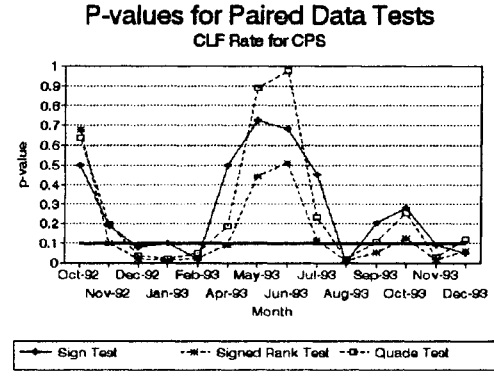

Figure 5

On the other hand, the paired data tests all reject the null hypothesis, with fairly small p-values. The p-value for the paired sign test using the fourteen month average is very small, and can even be rejected (at $\alpha = 0.05$) for a two-sided test. This conclusion is

neither proved nor disproved by the monthly p-value plot presented in Figure 5: this plot contains several large p-values for the paired sign tests. But obviously, this is not a very powerful test.

Both the fourteen month average and the monthly p-value plot for the Wilcoxon Signed Rank test provide evidence of a monthly effect for CLF participation rate. All of the assumptions (including symmetry) have been validated for this test, so the interpretation is straightforward. Again, the p-value is small enough that the test would be highly significant even for a two-sided test, as indeed it is in the Quade test.

B. Test For a New Questionnaire, Given Centralized Telephone Interviewing Effect Common CATI PSU's Data

1. Unemployment Rate

Table Three summarizes the nonparametric test results for a fourteen month average using unemployment rates.

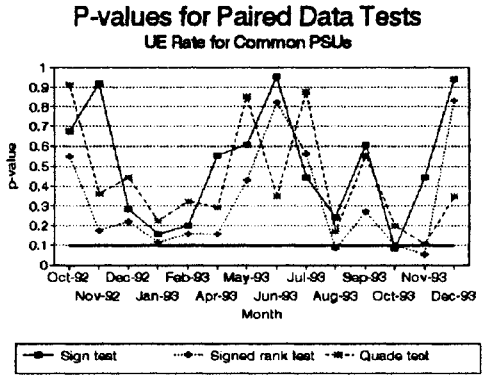

Figure 6

Neither of the two-sample tests provided evidence of this questionnaire effect for unemployment rate. On the surface, the paired data tests using fourteen month averages provide evidence of this effect. Further exploration does not reinforce this conclusion. No such trend is demonstrated in the monthly p-value plots for paired data tests presented in Figure 6. In fact, the plots show the reverse: in all three tests, the null hypothesis is rejected in **one** of fourteen months, fewer times than would be expected.

The "significant" paired data results for a fourteen month average are easily explained. The paired sign test result is unconvincing to begin with: this test would **not** reject for a two-sided test. The other paired data test results are explained by the effect of a small sample of PSUs described in section II.B.3. In this case, the eight smallest PSUs had the highest ranks in the fourteen month average.

Thus, we did not find any convincing evidence of this effect for unemployment rate.

2. Employment to Population Ratio

None of our tests found any evidence of a questionnaire effect for employment to population ratio.

3. CLF Participation Rate

None of our tests found any evidence of a questionnaire effect for CLF participation rate.

IV. Conclusion

Nonparametric analysis for this mode effects study provided new insights into the nature of the examined effects. The tests' results reinforced the published parametric CPS CATI Phase-in project results for unemployment rate, unencumbered by unprovable distributional assumptions. Moreover, the test results from the CPS split panel data provided reasonable evidence of a combined centralized and computer-assisted telephone interviewing effect for CLF participation rate.

The nonparametric analysis of CPS split panel data gave convincing results for two reasons:

1) test statistics were based on a large sample of PSUs;
2) panel estimates within a PSU had fairly balanced sample sizes.

Unfortunately, the Common CATI PSU analysis had neither a large sample of PSUs nor balanced sample sizes by panel within the PSU. Consequently, the nonparametric analysis failed to provide any more insight into a possible new questionnaire, given CATI effect on major labor force characteristics than the published parametric results provided.

V. Acknowledgements

We thank David Hubble, Alfredo Navarro, Norman Johnson, and Dedrick Owens for their useful comments. We also thank Mary Ann Braham for typing this paper.

VI. References

[1] Conover, W.J. Practical Nonparametric Statistics. New York: John Wiley and Sons, 1980.

[2] Hogg, Robert V. and Craig, Allen T. Introduction to Mathematical Statistics. New York: Macmillan Publishing Co., 1978.

[3] Memorandum for documentation from Jenny Seltzer and Robin Fisher, "CATI/CAPI Analysis: Averaging Factors for Nonparametric Analysis (CC_ALYS-20), July 12, 1993.

[4] Memorandum for Documentation from Robin Fisher and Richard McGuinness, "Correlations

and Adjustment Factors for CPS (VAR80-1)", January 6, 1993.

[5] Thompson, Jenny (1994), Mode Effects Analysis of Labor Force Estimates. *CPS Overlap Analysis Team Technical Report 3*, Bureau of the Census, Washington, D.C.

[6] Shoemaker, Harland (1993), Results from the Current Population Survey CATI Phase-in Project, *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

[7] Polivka, Anne E. (1994), Comparison of Labor Force Estimates for the Parallel Survey and the Current CPS, *CPS Overlap Analysis Team Technical Report 1*, Bureau of Labor Statistics, Washington, D.C.

This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

Table One: Unemployment Rate – CPS CATI Eligible PSUs 14 Month Average (10/92 through 12/93, excluding 3/93)

Test	Type of Test	P-Value
Two-Sample T-Test	Two-Sided	**0.0000** *
Mann-Whitney Wilcoxon	Two-Sided	**0.013** *
Paired Sign Test	One-Sided	**0.0001** *
Wilcoxon Signed Rank	One-Sided	**0.0000** *
Quade Test	Two-Sided	**0.0000** *

Table Two: CLF Participation Rate – CPS CATI Eligible PSUs 14 Month Average (10/92 through 12/93, excluding 3/93)

Test	Type of Test	P-Value
Two-Sample T-Test	Two-Sided	0.197
Mann-Whitney Wilcoxon	Two-Sided	0.167
Paired Sign Test	One-Sided	**0.016** *
Wilcoxon Signed Rank	One-Sided	**0.003** *
Quade Test	Two-Sided	**0.006** *

Table Three: Unemployment Rate – Common CATI PSUs 14 Month Average (10/92 through 12/93, excluding 3/93)

Test	Type of Test	P-Value
Two-Sample T-Test	Two-Sided	0.440
Mann-Whitney Wilcoxon	Two-Sided	0.354
Paired Sign Test	One-Sided	**0.064**
Wilcoxon Signed Rank	One-Sided	**0.038** *
Quade Test	Two-Sided	**0.075**