August 16, 1994

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

Jenny Thompson and Randall Parmer
Randall Parmer, Demographic Statistical Methods Division, Bureau of the Census, Washington, DC 20233

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

The analysis for mode effects in the CATI/CAPI Overlap Study has relied heavily on the two sample t-test. 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. Moreover, the variance estimate used in the parametric analysis is not distributed as a Chi-Squared random variable. Therefore, we apply a variety of nonparametric methods to split panel data from the Basic CPS and Parallel Survey, and show a comparison of the results to the normal theory based results.

1. 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. Moreover, the variance estimate used in the parametric analysis is not distributed as a Chi-Squared random variable.

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 more powerful than their parametric counterparts. A few examples:

1) In a two-sample analysis, the "Asymtotic Relative Efficiency (A.R.E.) of the Mann-Whitney test is never too bad when compared with the two-sample t test, the parametric counterpart. And yet the contrary is not true; the A.R.E of the t test compared to the Mann-Whitney test may be as small as zero, or 'infinitely bad'" ([1], p. 215);

2) The A.R.E of the Wilcoxon Signed Rank test or the Quade test relative to the paired data t test is "never less than 0.864" ([1], p. 291), under certain restrictions;

The nonparametric techniques are often effective even if the data are normally distributed. For normally distributed data, the A.R.E. of the nonparametric paired data techniques relative to the paired data t test is 0.955. Nothing is lost if the data are uniformly distributed: the comparative A.R.E is 1.0.

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. Appendix One contains descriptions of the four nonparametric tests and their applications. 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 the second appendix and [3]).

2. Hypotheses

2.1 Description of Split Panel Data

In both the CPS and the PS, the Census Bureau designated selected Primary Sampling Units (PSUs) as "CATI-eligible." 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 CATI interviewing 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, regardless of whether they met the above criteria.

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 pencil-and-paper questionnaire, and the computerized version of the questionnaire had a slightly modified wording of the lead-in question 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 pencil-and-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]. A direct comparison of the two surveys' designs is provided in *Appendix Three.*

2.2 Application of Nonparametric Tests to Split Panel Data

2.2.1 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.

Prior to applying any nonparametric tests to the split panel data, we checked 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 despite the following:

- Three of the CPS CATI Eligible PSUs had an average test panel sample size of less than 10 eligible persons;
- Two of the Common CATI Eligible PSUs had an average control panel sample size of less than 10 eligible persons.

Including these small PSUs in our analysis had potentially detrimental ramifications. Fortunately, in practice the small PSU/panel rates were almost invariably missing values in the tests of monthly data, and the PSUs' differences were not included in the tests. However, the effect of the small PSU/panel estimates did come into play when testing fourteen month averages. See section 2.2.3.3.

Generally, the number of CPS CATI Eligible PSUs was adequate for the analysis: seventy-five PSUs are included in the fourteen month average. The number CPS CATI Eligible PSUs each month ranged from a minimum of sixty-four PSUs to a maximum of seventy-two. 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.

2.2.2 Assumption Validation

Descriptions of each test, along with required assumptions, are provided in Appendix One.

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 used the simplified hypothesis specified in Appendix One.

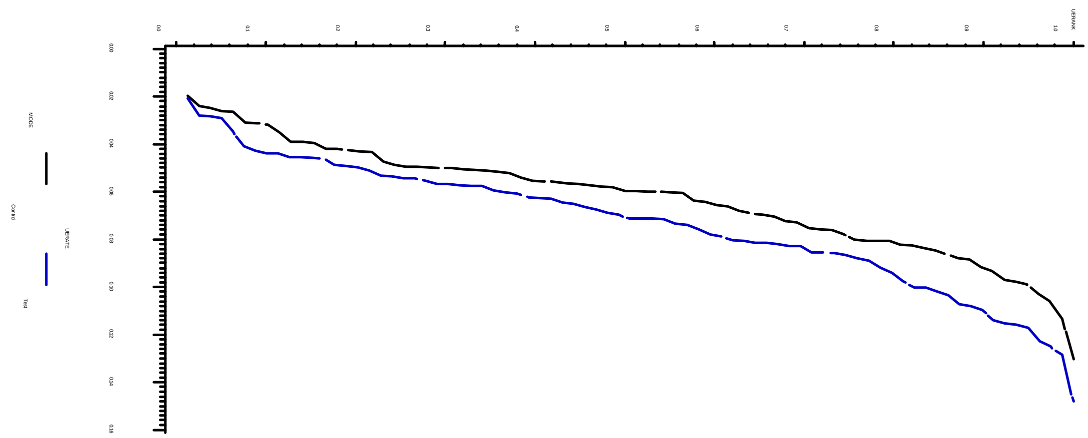

Figure 1: **CDF for Unemployment Rate - CPS CATI Eligible PSUs**

Empirical CDF plots for all Mann-Whitney tests are provided in Appendix Two.

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.

2.2.3 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 would be the ratio of the fourteen month averaged estimated unemployment level divided by the fourteen month averaged 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.

2.2.3.1 Background for Averaging Methodology

When originally planning the fourteen month nonparametric analysis, we were unsure whether we should use simple arithmetic means or incorporate previous knowledge of the statistics' autocorrelations into the fourteen month averages. To make a determination, we calculated quarterly, semi-annual, and annual estimates for four statistics using the two different methods, and then compared the variance obtained with these "optimal" factors to the variances obtained with simple arithmetic means.

The problem is best described by a series of equations. The linear program was

$$\min \operatorname{Var}\left(\sum a_i X_i\right)$$
$$\text{st } \sum a_i = 1$$

a_i are the averaging factors
x_i are the estimates for months i, i = 1...n
Var (X_i) = Var (X) for all i

In matrix form

Let r be the n x n correlation matrix for X̲. Assume that var(X_i) = var(x) for all X_i, so the correlation matrix is R, where the (i,j)th element is $\rho_{|i-j|}$.

Thus $Var(\sum a_i X_i)$ = except for a multiplicative constant (Var(X)) which is ignored without loss of generality.

The system to be solved is

to minimize, take the partial derivative with respect to a̲, and set equal to zero.

$$\frac{\partial\, b'R\,\underline{b}}{\partial \underline{a}} = 2J'RJ\,\underline{a} + 2J'R\underline{e_n}$$

$$\Rightarrow J'RJ\,\underline{a} = -J'R\,\underline{e_n}$$

$$\Rightarrow \underline{a} = -(J'RJ)^{-1}JR\,\underline{e_n}$$

Note that the second derivative is 2J'RJ which is positive definite for every value at a so this is the minimum. Note also that results obtained using the b defined by this a can be compared with those from the arithmetic mean.

2.2.3.2 Comparisons of Averaging Methodology

We used the replicate correlations from [4] for the R matrix to calculate averaging factors, variances using these factors, and variances calculated using an arithmetic mean for four statistics: total unemployed, total employed, total civilian labor force, and civilian non-institutional 16+ population. We compared the relative efficiency (in terms of variance) of using our "optimal" averaging factors to a simple arithmetic mean for three types of averages: quarterly, semi-annual, and annual.

The gain in variance reduction did not appear to offset the additional coding that these averaging factors would require. On the average, using optimal averaging factors decreased the variance by about one percent for quarterly estimates, four percent for semi-annual estimates, and one percent for annual estimates. Moreover, the nonparametric comparisons used rates rather than levels, and the "beneficial" effect of the averaging factors for a level might not perform in the same way for a rate.

Exact results are provided in [3].

2.2.3.3 Interpretation of Fourteen Month Averages

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 all 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, as in the case of the Common CATI Eligible PSU analysis, a few such PSU estimated differences may yield a "significant" result for a fourteen month average, even though the monthly results yielded consistently **non-significant** results.

2.2.4 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 Panel Employed	Test Panel Population	Test Rate	Control Panel Employed	Control Panel Population	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 (≈ .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.

2.2.5 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 characteristic estimates, and we therefore expected larger unemployment rates, employment to population ratios, and CLF Participation Rates for the treatment panels.

3. <u>Results</u>

Results are discussed by hypothesis. Boldfaced values are significant at α=0.10. An asterisk indicates the test is significant at α=0.05.

3.1 Tests for a Combined Centralized and Computer-Assisted Telephone Interviewing Effect -- CPS CATI-Eligible PSU Split Panel Data

3.1.1 Unemployment Rate

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

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	Two-Sample T-Test	Two-Sided	**0.0000 ***
Two Sample	Mann-Whitney Wilcoxon	Two-Sided	**0.013 ***
Paired Data	Paired Sign Test	One-Sided	**0.0001 ***
Paired Data	Wilcoxon Signed Rank	One-Sided	**0.0000 ***
Paired Data	Quade Test	Two-Sided	**0.0000 ***

As seen in Figure 1 (2.2.2), 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.

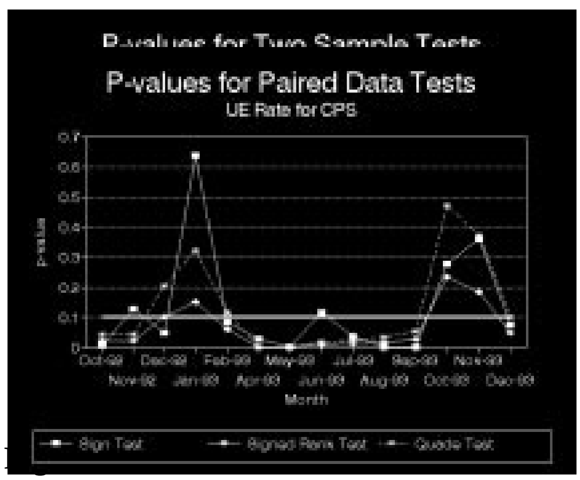

Figure 3

The results of the paired data tests provide more evidence for a combined centralization and computer-assisted interviewing 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.

3.1.2 Employment to Population Ratio

Table Two summarizes the nonparametric test results for employment to population ratio using fourteen month averages. These tests did not provide any evidence of a combined centralization and computer-assisted interviewing effect for this statistic.

Table Two: Employment to Population Ratio -- CPS CATI Eligible PSUs
14 Month Averag[illegible]/93)

	Test	[illegible]	P-Value
Two Sample	Two-Sample T-Test	[illegible]	0.984
Two Sample	Mann-Whitney Wilcoxo[illegible]	[illegible]	0.559
Paired Data	Paired Sign Test	[illegible]	0.455
Paired Data	Wilcoxon Signed Rank	[illegible]	0.133
Paired Data	Quade Test	[illegible]	0.268

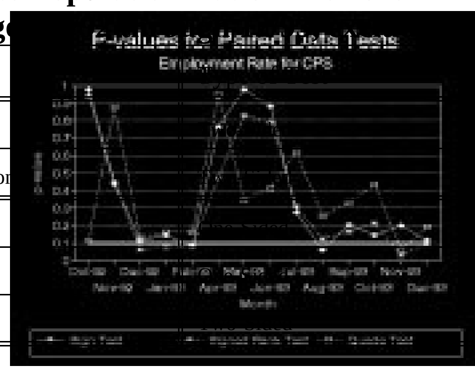

Figure 5

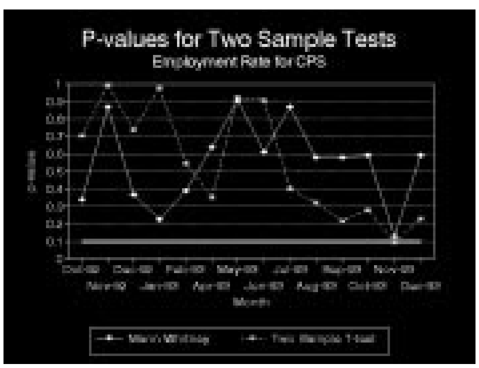

Figure 4

The monthly p-value plots presented in Figures 4 and 5 do not present any evidence of a combined centralized and computer assisted interviewing effect for monthly tests of employment to population ratios.

3.1.1 CLF Participation Rate

Table Three 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 for this hypothesis. 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 centralized and computer-assisted interviewing was included.

Table Three: 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	Two-Sample T-Test	Two-Sided	0.197
Two Sample	Mann-Whitney Wilcoxon	Two-Sided	0.167
Paired Data	Paired Sign Test	One-Sided	**0.016 ***
Paired Data	Wilcoxon Signed Rank	One-Sided	**0.003 ***
Paired Data	Quade Test	Two-Sided	**0.006 ***

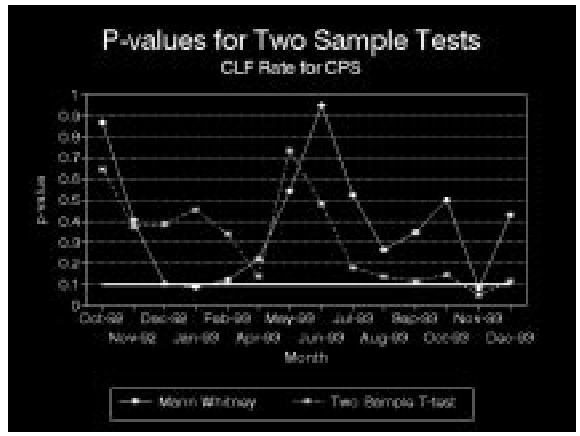

Figure 5

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. The CDF plots presented in Figure Three of Appendix Two reinforce the Mann-Whitney conclusion: if the two distributions are located differently, it is difficult to detect.

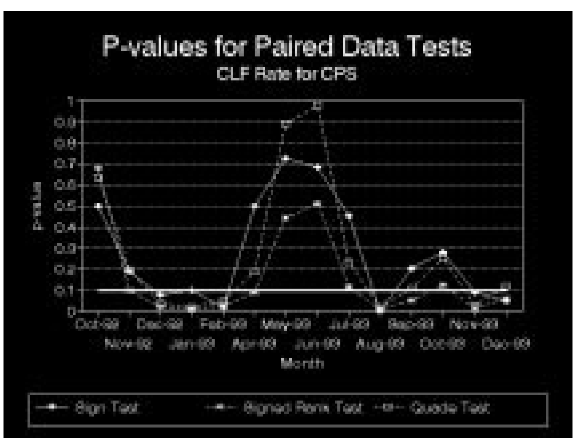

Figure 6

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 α=0.05) for a two-sided test. This conclusion is neither proved nor disproved by the monthly p-value plot presented in Figure 7: 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.

3.2 Test For a New Questionnaire, Given Centralized Telephone Interviewing Effect -- Common CATI PSUs' Data

3.2.1 Unemployment Rate

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

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

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

Neither of the two-sample tests provided evidence of a new questionnaire effect, given CATI, 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 8. 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.

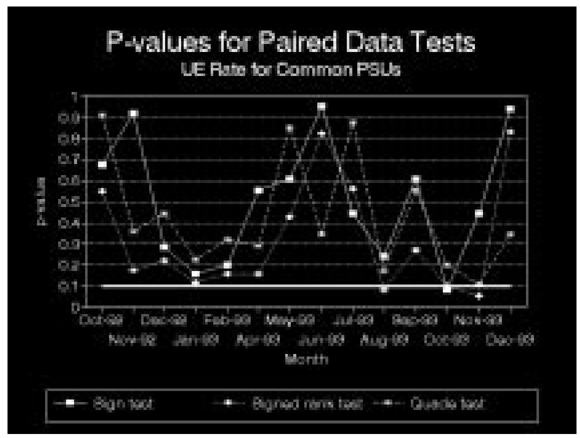

Figure 7

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 2.2.3.3. With a sample of 52 PSUs, the Wilcoxon Signed Rank test will reject a one-sided test if only the **eight** highest ranked differences are positive. In other words, this test will reject if ranks 45 through 52 are positive differences and if ranks 1 through 44 are non-positive. This event did not occur with the monthly data. However, with the fourteen month averages, the few small PSUs that had outlier positive difference are ranked along with the rest of the aggregated data, and the outcome is seen in Table Four.

Thus, we fail to find any convincing evidence of a new questionnaire, given centralized telephone interviewing, effect for unemployment rate.

3.2.2 Employment to Population Ratio

Table Five summarizes the nonparametric test results for a fourteen month average using employment to population ratio. These tests did not find any evidence of a new questionnaire, given a computer-assisted interviewing effect for this statistic. The monthly results did not give any indication of such an effect for this statistic.

Table Five: Employment to Population Ratio -- Common CATI PSUs
14 Month Average (10/92 through 12/93, excluding 3/93)

	Test	Type of Test	P-Value
Two Sample	Two-Sample T-Test	Two-Sided	0.839
Two Sample	Mann-Whitney Wilcoxon	Two-Sided	0.848
Paired Data	Paired Sign Test	One-Sided	0.661
Paired Data	Wilcoxon Signed Rank	One-Sided	0.717
Paired Data	Quade Test	Two-Sided	0.571

3.2.3 CLF Participation Rate

Table Six summarizes the nonparametric test results for a fourteen month average using the CLF participation rate. These tests did not find any evidence of a new questionnaire, given computer-assisted interviewing effect for this statistic. The monthly results did not give any indication of such an effect for this statistic.

Table Six: CLF Participation Rate -- Common PSUs
14 Month Average (10/92 through 12/93, excluding 3/93)

	Test	Type of Test	P-Value
Two Sample	Two-Sample T-Test	Two-Sided	0.730
Two Sample	Mann-Whitney Wilcoxon	Two-Sided	0.951
Paired Data	Paired Sign Test	One-Sided	0.445
Paired Data	Wilcoxon Signed Rank	One-Sided	0.590
Paired Data	Quade Test	Two-Sided	0.822

4. 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.

5. 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.

[8] Rothgeb, Jennifer M. (1994), Revision to the CPS Questionnaire Effect on Data Quality, *CPS Overlap Analysis Team Technical Report 2*, Bureau of the Census, Washington, D.C.

DSMD/CPS/JThompson/npar.doc

Appendix One
Nonparametric Tests

Mann Whitney Wilcoxon

References: [1], [2]

Purpose: To detect differences in the distribution functions based on independent random samples from two populations. Under a location shift alternative (see assumption 2), this test is the nonparametric analog to the two-sample t test.

Assumptions:
1. Independent Random Samples
2. The two CDFs are identically shaped. The difference between the two is due to a location shift. In other words, $F(X)$ is equal to $G(X+C)$, where C is a constant.

Hypothesis: If only the first assumption holds, use

(1) H_0: $F(X) = G(X)$
H_1: $F(X) \neq G(X)$

or equivalently,

H_0: $P(X < Y) = 1/2$
H_1: $P(X < Y) \neq 1/2$

If assumptions 1 and 2 hold, use

(2) H_0: $E(X) = E(Y)$
H_1: $E(X) \neq E(Y)$

Application: Each Mode Effect panel is a random sample of PSUs. The PSU/panel estimates are ranked within panel for analysis.

Paired Sign Test

References: [1],[2]

Purpose: To test whether one random variable in a pair (X,Y) tends to be larger than the other random variable in the pair.

Assumptions:

1. The bivariate random variables (X_i, Y_i), i=1,2,...,n are mutually independent (Note: X_i may or may not be independent of Y_i).

2. Each pair (X_i, Y_i) may be evaluated as a +, -, or tie.

3. The pairs (X_i, Y_i) are internally consistent.

Hypothesis:

H_0: $P(+) \leq P(-)$
H_1: $P(+) > P(-)$

or equivalently,

H_0: $E(X_i) \leq E(Y_i)$
H_1: $E(X_i) > E(Y_i)$

Application: Each PSU has a paired set -- Test Estimate and Control Estimate

Wilcoxon Signed Rank

References: [1],[2]

Purpose: To see if a "treatment" (such as CATI) results in the same median (uses paired data).

Assumptions:

1. The distribution function of the differences of the paired data is symmetric.

2. The pairs are mutually independent.

3. The pairs all have the same median.

4. (Optional) The (X_i, Y_i) for i=1,2,...n constitute a random (bivariate) sample.

Hypothesis: If only assumptions 1 through 3 hold, then

Let $d_{.50}$ = common median of the paired data differences

H_0: $d_{.50} \leq 0$
H_1: $d_{.50} > 0$

(i.e., the values of X tend to be smaller than the values of Y)

If all four assumptions hold, then the hypothesis is

H_0: $E(X) \leq E(Y)$
H_1: $E(X) > E(Y)$

Application: Each PSU has a random bivariate sample consisting of a test estimate and a control estimate. Panel differences for each PSU are ranked and analyzed.

Quade Test

Reference: [1]

Purpose: (Analogous to the parametric Two-Way ANOVA Model) To determine if a "treatment" (such as CATI) has an effect on the response variable, while blocking to alleviate the effect of the diversity of the sample units. This is a mathematical extension of the Wilcoxon Signed Rank Test.

Assumptions:

1. The treatments are independent within block.
2. The observations within block can be ranked.
3. The sample range for each block can be determined, so that the blocks themselves can be ranked.

Hypothesis:

H_0: Each ranking of the random variables within a block is equally likely (i.e., the treatments have identical effects).

H_1: At least one of the treatments tend to yield larger observed values than at least one other treatment.

Applications: Each PSU is a block. The PSU Test and Control Panel estimates are ranked within block, and the range (defined as the absolute value of the difference of the two) is calculated. The PSUs are then ranked based on this range.

Appendix Two
Cumulative Distribution Function Plots

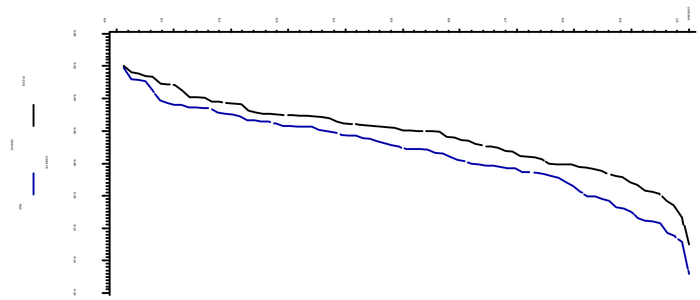

Figure One: CDF Plot for Unemployment Rate - CPS CATI Eligible PSUs

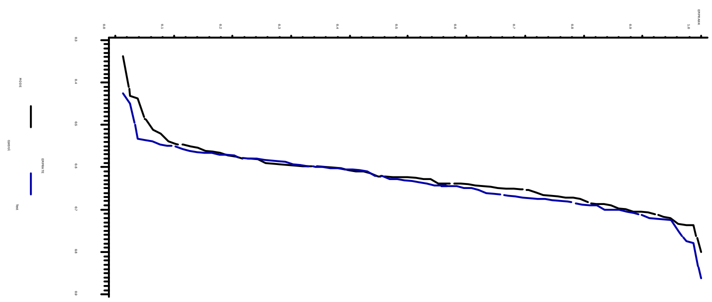

Figure Two: CDF Plot for Employment/Population - CPS CATI Eligible PSUs

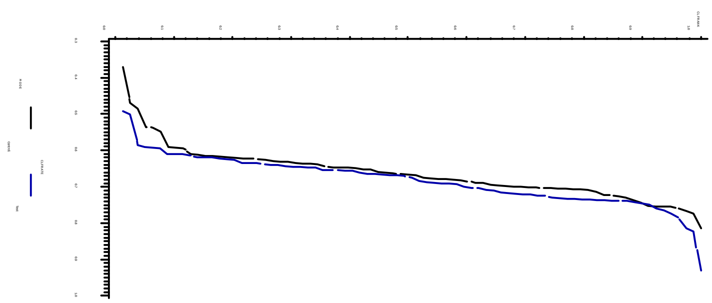

Figure Three: CDF Plot for CLF Participation Rate - CPS CATI Eligible PSUs

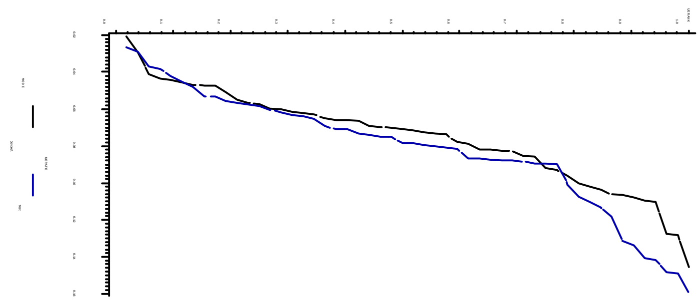

Figure Four: CDF Plot for Unemployment Rate - Common CATI PSUs

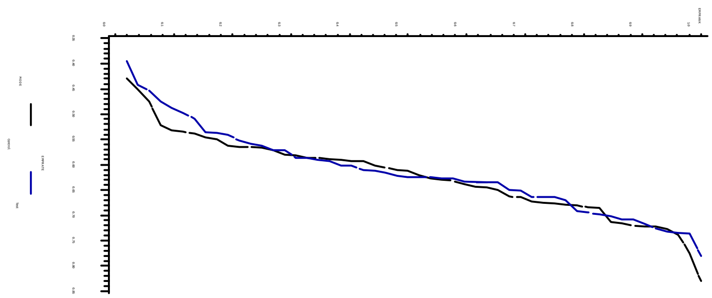

Figure Five: CDF Plot for Employment/Population - Common CATI PSUs

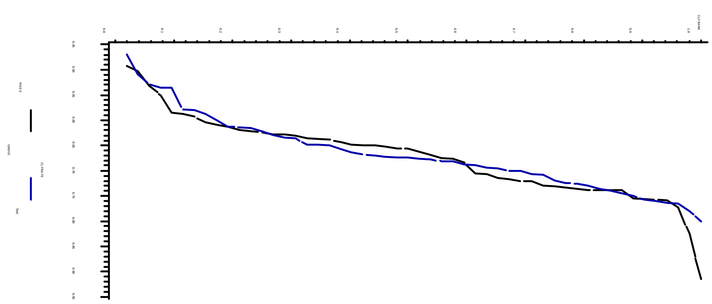

Figure Six: CDF Plot for CLF Participation Rate - Common CATI PSUs

BRIEF COMPARISON OF CPS AND PS DESIGNS

CPS DESIGN	PS DESIGN
Reliability and Sample Size	Reliability and Sample Size
1. 90% confidence interval on month to month change in unemployment rate of 0.2 percentage points.	1. 90% confidence interval on month to month change in unemployment rate of 0.4 percentage points.
2. Requirements for state estimates were: - 8% monthly CV or better for 11 largest states. - 8% annual CV or better for remaining states.	2. No state estimates or national reliability requirements. Design based on a fixed sample size.
3. Approximately 60,000 occupied housing units monthly.	3. Approximately 12,000 occupied housing units monthly.
4. 4-8-4 rotation pattern	4. 4-8-4 rotation pattern.
Design	Design
1. PSUs a. Defined as county or group of contiguous counties. b. PSUs correspond to projected 1983 MSAs. c. PSUs do not cross state boundaries. d. 729 sample PSUs; comprised of 1,297 geographic areas.	1. PSUs a. Majority have same definitions. Some CPS-level PSUs split up to hold down travel cost. b. PSUs usually correspond to projected 1983 MSAs. c. PSUs cross state lines, but not regional boundary. d. 283 sample PSUs; comprised of 579 geographic areas.
2. 1st-Stage (PSU) Selection a. NSR strata formed within a state. b. Used clustering algorithm to minimize variance; based on economic and labor force variables. c. PSU selection based on 1980 CNP 16+. d. 1 PSU per stratum selected.	2. 1st-Stage (PSU) Selection a. NSR Strata formed within a Census Region. b. Used clustering algorithm to minimize variance; based on demographic variables. c. PSU selection based on projected 1985 Housing units. d. 1 PSU per stratum selected.
3. 2nd-Stage of Selection a. EDs ordered within a PSU using	3. 2nd-Stage of Selection Same procedure as for CPS.

CPS DESIGN	PS DESIGN
geography and a clustering algorithm with economic variables. b. Segments of 4 housing units formed.	
Estimation and Weighting	Estimation and Weighting
1. 1st-Stage (PSU selection) adjusted for race within state.	1. 1st-Stage (PSU selection) adjusted for race within region.
2. Controlled to state totals in order to produce state estimates.	2. Not controlled to state totals, state-level estimation not practical.
3. Controlled to age/race/origin/sex totals at the national level.	3. Controlled to collapsed age/race/origin/sex totals at the national level to mimic CPS estimation at the national level.