

I. Introduction

The Consumer Price Index is based upon multiple samples, most of which are selected using complex multi-stage designs. The first stage of selection for these samples is a metropolitan or urban area. Within each selected area, a sample of outlets and a sample of items is selected for pricing each month. The focus of this paper will be the procedures used to select a new sample of areas or primary sampling units (PSU's) to be used in the revised Consumer Price Index, which is currently scheduled for release with January 1987 data.

Consumer Price Indexes are produced each month for two populations: all urban consumer units or the U population and urban wage earner and clerical worker consumer units or the W population. The W population is a subset of the U population. For each of the populations, indexes are produced for 28 cities and 12 region city-size areas at several levels of item aggregation.

Although the CPI uses complex estimators, it is fundamentally a weighted average of price changes from one time period to another for the same set of items. The weight for an item is an estimate of expenditures for the item at a particular point in time, the reference period. One of the major components in a CPI revision is the change in these reference period cost weights for each item-population-area which has a published index.

In order to obtain estimates of mean expenditures by item and geographic area for each index population, BLS conducts a Consumer Expenditure Survey (CES). For previous revisions, a one-time survey of households was conducted to obtain consumer expenditures. For the most recent 1972-73 CES, the Census Bureau acted as the data collection agency, while previously BLS had collected the data directly. Starting in January 1980, an ongoing rotating panel CES was put in place. Since the CES is also used to publish reports on consumption at the U.S. level¹, additional PSU's were selected to cover the rural non-metropolitan population. The CES is also used to select the item classes to be included in the revised index. The sample of outlets for the index is selected from a frame created by asking a different set of households the names and addresses of the store where they purchased categories of items. This Point of Purchase Survey (POPS) is also collected by the Census Bureau for BLS in the CPI PSU's.

II. PSU Definition

The current CPI/CES area sample was selected using 1970 Census population figures and 1973 Standard Metropolitan Statistical Area (SMSA) definitions. With the exceptions of the New York and Chicago Standard Consolidated Area's and the combination of Los Angeles-Long Beach SMSA and Anaheim-Santa Ana-Garden Grove SMSA, all metropolitan PSU's were defined as one SMSA. The new sample is based upon 1980 Census population and the new Consolidated Metropolitan Statistical Area (CMSA) and Metropolitan Statistical Area (MSA) definitions as proposed by the Office of Management and Budget. The change in definition from SMSA to CMSA is important. For instance, the definition for New York City now includes Danbury and other parts of Connecticut; Philadelphia includes Wilmington; Los Angeles includes Riverside-San Bernardino; and San Francisco includes San Jose.

All MSA's which are not a part of a CMSA are defined as individual PSU's. All remaining nonmetropolitan counties were grouped into PSU's using the following criteria:

1. To the extent possible, the old PSU definitions were to be retained.
2. The counties forming a PSU had to be contiguous.
3. The number of urban housing units in the PSU had to be at least 3125, the number of housing units required to support the CES and POPS for 10 years.

The same set of counties were then regrouped into another set of PSU's to be used to represent the rural nonmetropolitan portions of the country. Since the old urban and rural PSU's were created independently, any concordance between urban and rural PSU composition was happenstance.

A total of 2472 PSU's were created: 278 metropolitan areas, 810 nonmetropolitan urban areas, and 1384 rural areas.

III. Overall Sample Design

The basic area sample design plan for the CPI consisted of the selection of one PSU per stratum with probability proportional to population. Within this frame work, a number of constraints were imposed on the allocation of the sample.

1. The proposed ongoing CPI budget would permit the pricing of $127\frac{1}{2}$ halfsample equivalents² every two months. A halfsample consists of approximately 1100 quotes or item-outlet pricings. The current CPI has 119 halfsample equivalents or 120,000 unique prices are collected within each two month period. Thus the revised CPI will have about a 6.7% increase in unique quotes.
2. The number of areas should be around $90\frac{2}{3}$. The CPI currently collects data in 85 areas.
3. The option to continue publication of all city and region city-size indexes currently being published with the exception of Northeast Pennsylvania should be allowed for. In order to insure a city is in the sample, it must be self-representing.
4. Although the general scheme for city-size classification should be maintained, the specifics could be modified. In the 1970 design, all nonself-representing SMSA's were subdivided into two classes: those greater than 400,000 and those less than 400,000. The nonmetropolitan urban areas were stratified independently.
5. To the extent possible, the sample should be allocated proportionally among the region and city-size classes.
6. In order to publish a region₄ city-size class, there must be at least four⁴ sample PSU's within the area implying at least eight PSU's are needed in the frame population; otherwise, at least one PSU would end up self-representing.

The total population of the U.S. to be covered by the revised CPI (the CPI U population) is 187,086,922 as of April 1, 1980. In a sample design with proportional allocation, any PSU with a population of more than $187,086,922/127=1,473,125$ should be self-representing.

Therefore, the 23 largest cities become self-representing with Phoenix being the last self-representing MSA. These cities have a CPI population of 91,410,272 or 49% of the total. The remaining 75,592,095 metropolitan CPI population is evenly divided between the medium and small sized MSA's.

When this proportional design is examined with respect to criteria three and six, problems appear. To begin with, several cities currently being published are not among the largest 23 cities--Kansas City, Milwaukee, Portland, Buffalo, Honolulu and Anchorage. Secondly, there are only six PSU's classified as medium sized cities in the West. Since four sample PSU's are required for publication, two of the six PSU's would be self-representing. Therefore additional cities must be classified into this region city-size class.

After making the six published cities self-representing, numerous allocations between medium and small sized cities indicated no consistent breaking point across regions yielded a reasonable allocation. As a result, each region was treated individually and the breaking points became 500,000 in the Northeast, 360,000 in the Northcentral, 450,000 in the South and 330,000 in the West. Table 1 compares the number of PSU's and the distribution of the CPI population in the proportional allocation scheme to that in the final allocation scheme. In this final design, 53.5% of the CPI population or 45% of the total is in self-representing PSU's.

The next step in the design was to determine the number of strata to be created in each region city-size class. If four strata were created in the nonmetropolitan urban Northeast, the average strata size would be 500,000 implying 174 strata would be needed in nonself-representing areas alone in order to maintain equal sized nonself-representing strata. Since this was impossible, numerous alternatives were examined. In the end, the decision was made to drop the publication requirement in the nonmetropolitan urban areas for the Northeast and West. Table 2 indicates the final sample PSU allocation scheme of 109 sample PSU's for the CES and 91 sample PSU's for the CPI and POPS. In addition, the allocation scheme of the current area sample is included for comparison purposes.

In the final design, 46.5% of the CPI population is in nonself-representing PSU's. The NSR strata are as equal as possible given the constraint of an even number of PSU's in each region city-size class. Therefore, within the nonself-representing portion of the design the CPI sample is as proportionally allocated to population as possible.

If the current CPI publication requirement of publishing three consistently defined city-size classes per region were continued for the revised sample, the variance of the national CPI would be 15% higher. By allowing the city-size break to be variable across region and not publishing (selecting four PSU's in a size class) the nonmetropolitan areas of the Northeast and West, this larger variance has been avoided.

IV. Stratification

A. Procedure

The primary objective of stratification is to create strata as homogeneous as possible in order to minimize the between PSU component of variance within a set of constraints on the size of the clusters or strata.

The between PSU variance function to be minimized is:

$$\sum_{j=1}^k f_j \sum_{h=1}^g \sum_{i=1}^{n_h} \frac{P_{hi}}{P_h} \left(\frac{Y_h}{Y_{hi}} U_{hij} - U_{hj} \right)^2, \quad (1)$$

where
 g = the number of strata in the region city-size class,
 n_h = the number of PSU's in the hth stratum,
 k = the number of variables in which the stratification is to be based,
 f_j = the preference factor or weight for the jth characteristic,
 P_{hi} = the population of the ith PSU in the hth stratum,
 P_h = the population of the hth stratum,
 Y_{hi} = the number of housing units in the ith PSU in the hth stratum,
 Y_h = the number of housing units in the hth stratum,
 U_{hij} = the number of housing units in the ith PSU in the hth stratum having characteristic j or the number of dollars spent for item j,
and
 U_{hj} = the number of housing units in the hth stratum having characteristic j or the number of dollars spent for item j.

Letting $X_{hij} = \frac{U_{hij}}{Y_{hi}}$ and $X_{hj} = \frac{U_{hj}}{Y_h}$,

the proportion of housing units in PSU i or stratum h having characteristic j or the average housing unit expenditure or income, the above formula can be rewritten as

$$\sum_{j=1}^k f_j \sum_{h=1}^g \frac{Y_h^2}{P_h} \sum_{i=1}^{n_h} P_{hi} (X_{hij} - X_{hj})^2. \quad (2)$$

B. Implementation

The process of creating strata from the PSU's within a region city-size class was implemented using a program developed by the Census Bureau which used Friedman-Rubin's clustering algorithm (1) to stratify PSU's within a state for the Current Population Survey (CPS). Documentation of the Census Bureau's research into clustering and implementation of the Friedman-Rubin clustering algorithm can be found in (2).

The basic purpose of cluster analysis is the identification of the "best partition" of a set of objects into 'g' groups where "best partition" is defined by optimizing some numerical valued function. In this case, the 'n' PSU's are grouped into 'g' strata by minimizing equation (2) above. The modified Friedman-Rubin algorithm successively reallocates PSU's to strata, using several different procedures to determine which moves between strata result in a reduction in between PSU variance. Since the algorithm seeks to modify some initial stratification in ways which reduce the between PSU variance, there is no guarantee an absolute minimum is achieved.

Numerous modifications had to be made to the program in order to implement its use for the CPI. Significant changes were necessary to the variance computation portions of the program since the characteristics used for the CPI stratification were in terms of housing units versus population for CPS.

The basic input to the program consisted of the number of strata required, the limits on the strata sizes, the variables to be used in the clustering and the preference factors for each of the characteristics.

1. Selection of Stratification Variables

Since the primary purpose of the CPI is to measure price change, the characteristics used to stratify PSU's for the CPI should be highly correlated with price change. From the Census Bureau, we obtained a special file of over 200 population and housing unit charac-

teristics by county (MCD in the Northeast) from the 1980 Census which we entered into a SAS database. A subset of 21 variables (Table 3) was created and merged with one year changes (1980) and four year changes (1978 through 1982) of the CPI for each of ten major groups from the 27 self-representing areas of the current CPI. Multiple linear regressions were run to determine which variables were associated with price change and correlation matrices were computed to determine the correlation among the variables. The seven variables asterisked in Table 3 were chosen to be used in the stratification program.

The R^2 for the all items index using an estimated 'rental equivalence' index (3) for housing with these variables was .67, a level which exceeded our expectations.

2. Selection of Preference Factors

Several sets of preference factors were examined to determine what effects the choice of preference factors had on the stratification process and the resulting variance. In the end, two sets of preference factors were used for each region city-size class:

| Variable | Preference Factors | |
|-----------------------------------|--------------------|---------|
| Owner Occupied | .5 | .10373 |
| Retired Reference Person | .5 | .22680 |
| Black Reference Person | .5 | .43499 |
| Mean Wage and Salary | 1.0 | .51472 |
| Heated by Fuel Oil | .5 | .81253 |
| Heated by Electricity | .5 | .91775 |
| Income | | |
| Mean Interest and Dividend Income | 1.0 | 1.00000 |

The first set of factors places the greatest importance on the two income variables without differentiating among the non-income variables. Since income is highly correlated with expenditures, these factors were assumed to be important for controlling the between PSU component of variance for the CES.

The second set of factors were based upon the relative correlation of each of the variables with price change where interest and dividend income was the variable most highly correlated. This set of factors was assumed to be reasonable for controlling the between PSU component of variance for the CPI.

C. Results

Since the results of the algorithm depend upon the initial stratification, the program allows for multiple starts. Each region city-size class was stratified at least ten times using each set of preference factors. For the medium sized MSA's it was possible to actually look at the physical composition of the strata since there were only two or three PSU's per strata. For instance, in the Northeast only four different stratifications appeared meeting the size constraints. As the number of PSU's increased it became impossible to examine the stratifications in any way other than some set of summary statistics.

In order to help differentiate among the 20+ stratifications, the resulting between PSU variances for each of the seven stratification variables along with four additional variables (average monthly expenditures for gas and electricity and percent of U and W HU's) were ranked along with the range of the strata sizes. The selection of the stratification to be used was based upon the sum of the ranks, the total between PSU variances for the seven stratification variables both unweighted and weighted⁵ by the preference factors, the strata size range, and the placement of the current sample PSU's. The last two criteria were considered

the most important given the selection was among those stratifications which minimized the variance given the initial stratification and preference factors. The range of the strata size was considered important since the NSR PSU's are paired and the PSU's in a pair are priced alternately, i.e., bimonthly. Therefore, each PSU in a pair should represent equivalent populations. The placement of the current sample PSU's was considered important since the new sample is to be introduced over a period of several years. Any new stratum not containing an old sample PSU would have to be initiated prior to January 1987; while if a new stratum contained an old sample PSU, pricing could continue in the old PSU until the new sample PSU is initiated.

V. Computation of Probabilities of Selection

Since the introduction of new PSU's requires hiring and training new field staff, the costs associated with changing PSU's are not insignificant; therefore, there is some desire to retain as many old PSU's as possible, while at the same time incorporating the new population data. On the other hand, there is a concern over respondent burden. Although the basic probability of selection for the i th PSU from the h th stratum is P_{hi}/P_h , Keyfitz (4) developed a procedure for adjusting these probabilities so that the probability of retaining the old sample PSU is maximized given the old stratification is retained.

Since Keyfitz's initial article, several authors have developed similar procedures for the situation where the strata are changed. We considered two of these procedures: one developed by Walter Perkins for use in the 1970 CPS redesign (5), and the second by Lawrence Ernst for use in the 1980 CPS redesign (6). The Perkins procedure is not optimal with respect to retaining current sample PSU's while the Ernst method is.

We experimented with both procedures by running the medium sized MSA strata in the Northeast and West and the small MSA strata in the West for a total of 11 strata. Table 4 shows the 1970 and 1980 basic probabilities along with the 1980 Ernst and Perkins probabilities. The column labelled sample contains a '1' if the PSU is currently in sample. Under Perkins' method, if there is only one PSU from a 1970 stratum in the new 1980 stratum, the probability of selecting that PSU remains unchanged. Therefore in the medium sized strata where none of the 1970 pairings were retained, Perkins method leaves the probabilities unchanged. On the other hand, Ernst's method changes the probabilities dramatically and, in fact, virtually selects the sample. Of the four small MSA strata, two contain a current sample PSU and two do not. Ernst selects a PSU in one of the sample cases but leaves the other sample case virtually untouched. Some of the changes in the strata without a sample PSU are large under the Ernst method while Perkins' leaves them untouched.

In the end, Perkins method was used mostly because we felt Ernst's method was too extreme. Our goal in adjusting the probabilities was to increase overlap over an independent selection while at the same time reflecting the shifts in population between Censuses and avoiding undue respondent burden. Since the CPI is basically a longitudinal survey where the selected retail outlets will be visited every one or two months for 5 or 10 years, there are respondent burden problems associated with retaining smaller PSU's. The subtotal of NSR probabilities for the overlap PSU's increased 73% from 12.66 for the 1980 probabilities to 22.00 for the Perkins' probabilities. Or, of the total NSR probability (78.0), the overlap probability increased from 16.2% to 28.2%.

VI. Controlled Selection

In order to insure the representation of the sample by state is directly proportional to the population of each state, a three-way controlled selection procedure was used. The algorithm developed for the 1970 CPS redesign (6) was modified and executed by Census region controlling on state, strata and overlap.

Basically, if the sum of the NSR PSU probabilities associated with a state is 2.59 then the constraint of controlled selection is to select either two or three PSU's from that state. The process consists of creating many possible samples or patterns which meet the constraints and have a known probability of selection, such that the sum of the pattern probabilities equals 1. Across all patterns, a cumulative state probability of 2.59 should result in two PSU's from the state in 41% of the patterns and three PSU's in 59% of the patterns. Controlled selection allows preferred combinations of PSU's to be selected while maintaining the basic probabilities of selection.

Since the primary control is on state and overlap is secondary, the overlap/nonoverlap proportions of the sample are not exact. That is, although only 22 overlap PSU's should have been selected with probability .9971, 24 were selected. Other cases which differ from expectation are: Florida - nonoverlap which expected two or three has four sample PSU's, Massachusetts-overlap which expected zero or one has two, Ohio-nonoverlap which expected two or three has four, Ohio-overlap which expected one or two has zero, and Utah-nonoverlap which expected zero or one has two.

On an overall basis, including the SR PSU's, the breakdown of the sample between new and old PSU's by region is as follows.

| Region | Number of New and Old PSU's | | Number of New and Old PSU's | |
|--------|-----------------------------|-------|-----------------------------|-------|
| | C E S | C P I | C E S | C P I |
| NE | 6 | 11 | 6 | 9 |
| NC | 14 | 15 | 8 | 15 |
| S | 26 | 16 | 18 | 16 |
| W | 9 | 12 | 7 | 12 |
| Total | 55 | 54 | 39 | 52 |

Although in terms of PSU's at the U.S. level the CES sample is about half new and half old, almost half of the new PSU's are located in the South. For the CPI, over half of the current PSU's are being retained, but again almost half of the new PSU's are in the South. In terms of population, 57.4% of the CPI urban population is located in overlap PSU's and of the CPI urban population in the new sample PSU's, 79.6% live in counties included in the current sample.

VII. Sample Introduction

The new CPI PSU's will be introduced over a two year period even though the revised index is released in February 1987. In order to do this, some of the old PSU's will be used to represent their new strata. Any new strata which does not contain a current sample PSU must have its sample PSU initiated before January 1987 so it can be linked into the index in January. There are a total of 20 such strata, one SR (Phoenix) and 19 NSR. These 20 cities will be in the POPS survey of 1985, initiated by BLS beginning in May 1986 and then linked into the index with January 1987 data. Ten of the remaining 19 new PSU's will be linked during 1987 with the remaining nine in 1988.

VIII. Further Research and Summary

In summary, there are several areas where further research is warranted.

1. The significant costs associated with stratification coupled with the facts the between PSU variance component is small (5 to 10%) and there are no Census data directly associated with changes in consumer prices, raises the question as to whether such a sophisticated algorithm should be used for stratification in the future for this size sample. A simpler algorithm might be developed which would be more cost effective with respect to reducing the between PSU variance.

2. Considering the significant changes in probability under Ernst procedure for the few strata examined, further research is needed to determine how the procedure works under varying conditions and if additional constraints could be added to control the percent change in probability.

3. With the use of controlled selection, the selected PSU's are no longer independent of each other. The variance of the controlled selection procedure is unknown and should be examined along with determining what, if any, effects it has on the estimation of variances using pseudo-replications.

*The authors wish to acknowledge the efforts of Mary French and William McCarthy of SMD/BLS who were responsible for executing the controlled selection process defined in this paper.

- 1/ Expenditures at the U.S. level will not be published until the release of the 1984 CES data.
- 2/ The revision budget proposals were developed using three criteria: 1. The new design should reflect the population shift to the South and West, 2. the between PSU variance should be no larger than that for the 1978 revision, 3. increases should be kept to a minimum.
- 3/ The CPI sample is designed to facilitate computation of variances using pseudo replication. Within each SR area two halvesamples are designated. In order to increase the sample in an area such as Chicago to the equivalent of three halvesamples, two halvesamples, each with 1½ times the number of quotes in other SR halvesamples, would be designated. A single halvesample is designated in each NSR PSU, and the NSR PSU's are paired for variance computation.
- 4/ Four sample PSU's are needed to allow for bimonthly pricing and computation of variances using pseudo replications.
- 5/ The variances are standardized at the beginning of the program; therefore, the weighted sum of variances is equivalent to the weighted sum of relvariances.

References

1. Friedman, H.P. and Rubin, J. "On some Invariant Criteria for Grouping Data": Journal of the American Statistical Association 62, 1159-1179, 1967.
2. Kostanich, D., Judkins, D., Singh, R, and Schautz, M. "Modification of Friedman-Rubin's Clustering Algorithm for Use in Stratified PPS Sampling". Proceedings of the American Statistical Association, August 1981.
3. Gillingham, Robert and Lane, Walter. "Changing the Treatment of Shelter Costs for Homeowners in the CPI". Monthly Labor Review, BLS, U.S. Department of Labor, 1982, June 9-14, Vol. 105.
4. Keyfitz, Nathan. "Sampling with Probabilities Proportional to Size: Adjustment for Changes in the Probabilities" Journal of the American Statistical Association 46, 105-109, 1951.
5. U.S. Bureau of the Census. The Current Population Survey - Design and Methodology. Technical Paper 40. 1978.
6. Ernst, Lawrence R. Optimal Methods for Maximizing and Minimizing Overlap Between Surveys. Unpublished memorandum.

T A B L E 1

COMPARISON OF PROPORTIONAL AND FINAL SAMPLE DESIGNS

| | <u>PROPORTIONAL</u> | | | | | <u>FINAL</u> | | | | |
|--------|----------------------------------|-----------|----------|----------|--------------|--------------|-----------|----------|----------|--------------|
| | <u>CPI POPULATION (MILLIONS)</u> | | | | | | | | | |
| | <u>NE</u> | <u>NC</u> | <u>S</u> | <u>W</u> | <u>Total</u> | <u>NE</u> | <u>NC</u> | <u>S</u> | <u>W</u> | <u>Total</u> |
| SR | 29.1 | 21.2 | 17.6 | 23.5 | 91.4 | 30.4 | 25.1 | 18.9 | 25.6 | 100.0 |
| Medium | 8.3 | 8.4 | 15.8 | 5.0 | 37.5 | 6.6 | 6.0 | 15.1 | 4.6 | 32.2 |
| Small | 5.2 | 10.6 | 15.9 | 6.3 | 38.1 | 5.7 | 9.0 | 15.4 | 4.6 | 34.8 |
| Urban | 2.0 | 6.4 | 8.1 | 3.6 | 20.1 | 2.0 | 6.4 | 8.1 | 3.6 | 20.1 |
| Rural | 4.0 | 10.7 | 15.8 | 3.8 | 34.4 | 4.0 | 10.7 | 15.8 | 3.8 | 34.4 |
| Total | 48.6 | 57.3 | 73.3 | 42.2 | 221.5 | 48.6 | 57.3 | 73.3 | 42.2 | 221.5 |

ALL PSU'S

| | <u>NE</u> | <u>NC</u> | <u>S</u> | <u>W</u> | <u>Total</u> | <u>NE</u> | <u>NC</u> | <u>S</u> | <u>W</u> | <u>Total</u> |
|--------|-----------|-----------|----------|----------|--------------|-----------|-----------|----------|----------|--------------|
| SR | 4 | 6 | 7 | 6 | 23 | 5 | 9 | 8 | 9 | 31 |
| Medium | 11 | 9 | 21 | 6 | 47 | 9 | 10 | 21 | 8 | 48 |
| Small | 27 | 61 | 88 | 32 | 208 | 28 | 57 | 87 | 27 | 199 |
| Urban | 69 | 246 | 372 | 123 | 810 | 69 | 246 | 372 | 123 | 810 |
| Rural | 108 | 479 | 618 | 179 | 1384 | 108 | 479 | 618 | 179 | 1384 |
| Total | 219 | 801 | 1106 | 346 | 2472 | 219 | 801 | 1106 | 346 | 2472 |

T A B L E 2

ALLOCATION OF SAMPLE PSU'S

| | <u>Current</u> | | | | | <u>New</u> | | | | |
|--------------|----------------|-----------|----------|----------|--------------|------------|-----------|----------|----------|--------------|
| | <u>NE</u> | <u>NC</u> | <u>S</u> | <u>W</u> | <u>Total</u> | <u>NE</u> | <u>NC</u> | <u>S</u> | <u>W</u> | <u>Total</u> |
| SR | 6 | 8 | 6 | 7 | 27 | 5 | 9 | 8 | 9 | 31 |
| Medium | 4 | 4 | 8 | 4 | 20 | 4 | 4 | 10 | 4 | 22 |
| Small | 4 | 6 | 8 | 4 | 22 | 4 | 6 | 10 | 4 | 24 |
| Non-MSA | 4 | 4 | 4 | 4 | 16 | 2 | 4 | 6 | 2 | 14 |
| CPI Subtotal | 18 | 22 | 26 | 19 | 85 | 15 | 23 | 34 | 19 | 91 |
| CES Rural | 2 | 4 | 8 | 2 | 16 | 2 | 6 | 8 | 2 | 18 |
| Total | 20 | 26 | 34 | 21 | 101 | 17 | 29 | 42 | 21 | 109 |

T A B L E 3

VARIABLES REGRESSED AGAINST PRICE CHANGE

Description

- *Percent black HU's
- Percent two or more earner HU's
- *Percent electric heated HU's
- Average family size
- Percent natural gas heated HU's
- *Percent owner occupied HU's
- Average income per person
- Mean monthly electric costs per HU
- Mean total income per HU
- *Mean interest & dividend income per HU
- Mean monthly gas costs per HU
- Mean yearly oil costs per HU
- *Mean wage and salary income per HU
- Mean real estate taxes per HU
- Median value of owned home
- Median contract rent
- *Percent fuel oil heated HU's
- *Percent HU's with retired person
- Percent unemployed persons
- Percent CPI urban HU's
- Percent wage & clerical HU's
- *Selected variables

Table 4 - Comparison of Ernst and Perkins Probabilities for Increasing Overlap

| PSU | Sample | <u>1970 Probability</u> | <u>1980 Probability</u> | <u>Ernst</u> | <u>Perkins</u> |
|-------|--------|-----------------------------|-----------------------------|--------------|----------------|
| B1.11 | 1 | .57 | .62 | 1.00 | .62 |
| B1.12 | 0 | .39 | .38 | 0.00 | .38 |
| B1.21 | 1 | .51 | .68 | .04 | .68 |
| B1.22 | 1 | .31 | .32 | .96 | .32 |
| B1.31 | 1 | 1.00 | .36 | 0.00 | .36 |
| B1.32 | 0 | .25 | .30 | .05 | 0.00 |
| B1.33 | 1 | .36 | .35 | .95 | .65 |
| B4.11 | 0 | .41 | .76 | .99 | .76 |
| B4.12 | 0 | .23 | .24 | .01 | .24 |
| B4.21 | 0 | .47 | .70 | 0.00 | .70 |
| B4.22 | 1 | .27 | .30 | 1.00 | .30 |
| B4.31 | 1 | .30 | .56 | 1.00 | .56 |
| B4.32 | 0 | .26 | .44 | 0.00 | .44 |
| B4.41 | 1 | .26 | .51 | 1.00 | .51 |
| B4.42 | 0 | .22 | .49 | 0.00 | .49 |
| C4.11 | 0 | .21 | .25 | 0.00 | .25 |
| C4.12 | 0 | .14 | .22 | 0.00 | 0.00 |
| C4.13 | 1 | .19 | .23 | 1.00 | .45 |
| C4.14 | 0 | .08 | .12 | 0.00 | .12 |
| C4.15 | 0 | .05 | .10 | 0.00 | .10 |
| C4.16 | 0 | .07 | .08 | 0.00 | .08 |
| C4.21 | 1 | .19 | .26 | .33 | .26 |
| C4.22 | 0 | .22 | .28 | .12 | .28 |
| C4.23 | 0 | .10 | .11 | .14 | .11 |
| C4.24 | 0 | .07 | .07 | .08 | .07 |
| C4.25 | 0 | .13 | .20 | .23 | .20 |
| C4.26 | 0 | .06 | .08 | .10 | .08 |
| C4.31 | 0 | .17 | .23 | .09 | .23 |
| C4.32 | 0 | .15 | .20 | .08 | .20 |
| C4.33 | 0 | .11 | .14 | .20 | .14 |
| C4.34 | 0 | .05 | .11 | .16 | .11 |
| C4.35 | 0 | .05 | .09 | .13 | .09 |
| C4.36 | 0 | .06 | .11 | .16 | .11 |
| C4.37 | 0 | .05 | .12 | .18 | .12 |
| C4.41 | 0 | .10 | .17 | .22 | .17 |
| C4.42 | 0 | .11 | .19 | .24 | .19 |
| C4.43 | 0 | .09 | .14 | .07 | .14 |
| C4.44 | 0 | .07 | .09 | .03 | .09 |
| C4.45 | 0 | .07 | .12 | .06 | .12 |
| C4.46 | 0 | .08 | .13 | .17 | .13 |
| C4.47 | 0 | .05 | .10 | .13 | .10 |
| C4.48 | 0 | .05 | .06 | .08 | .06 |