

SAMPLE DESIGN(S) FOR THE CONSUMER PRICE INDEX REVISION (1978)

Curtis A. Jacobs, Bureau of Labor Statistics

I. INTRODUCTION

The purpose of this paper is to document the sample design considerations which are used in the development of the estimates of the price relatives for the Consumer Price Index Revision.¹

This paper describes the general structure and sampling procedures to select the items, outlets and specifications within outlets and the general estimation techniques for the two family indexes for all components of the CPIR.

II. DESCRIPTION OF DATA SOURCES

A. CES

The source of the detailed expenditures is the 1972-73 Consumer Expenditure Survey. The 1972-73 Consumer Expenditure Survey consisted of two separate components, each with its own questionnaire and sample: (1) An interview panel survey in which each consumer unit in the sample was visited by an interviewer every 3 months over a 15-month period; and (2) a diary or recordkeeping survey completed by respondents for two consecutive 1-week periods.

1. Diary Survey

The diary survey was divided into two 12-month periods. The primary objective of the survey was to obtain expenditure data on small, frequently purchased items which are normally difficult to recall. These items include expenditures for food and beverages, gas and electricity, gasoline, housekeeping supplies, nonprescription drugs and medical supplies, and personal care products and services. The daily record was designed as a self-reporting, product-oriented diary on which respondents recorded all expenses for two consecutive 1-week periods.

2. Quarterly Interview Survey

The quarterly interview survey was designed to collect data on major items of expense as well as on income and family characteristics. Panels of sample households were established to allow for the collection of data throughout each quarter, with each sample unit being interviewed once per quarter for five consecutive quarters.

Interviews in the first quarter provided socioeconomic characteristics of the consumer unit and an inventory of durable items owned by the CU. The inventory items were recorded to prevent the duplication of expenditures in subsequent interviews. The first-quarter interview also recorded data covering a great variety of regularly-purchased items bought since the first of the year. The recall period for reporting data varied according to the difficulty of recall for a class of items. Housing, major equipment, vehicles, subscriptions and insurance were collected annually. A semi-annual recall period was used for minor equipment, housefurnishings, renting and leasing of vehicles, and education. The following items were covered each quarter: repairs, alterations and maintenance of owned

property; utilities, fuels, and household help; clothing and household textiles; equipment repairs; vehicle operating expenses; and out-of-town trips. The final interview in the fifth quarter provided the regularly-recorded expenses, plus information on housing expenses (i.e., ownership costs); work experience, changes in assets and liabilities, estimates of consumer unit income, and other selected financial information.

3. Sample selection for CES

For both surveys, the nation was divided into 216 geographic areas using stratification variables defined for the Current Population Survey. Thirty of these areas were self-representing, i.e., selected with certainty, primarily because of the size of the population areas represented. Also twenty-seven of these areas corresponded to CPIR required local publication areas. Half of the housing units in each of these self-representing areas was covered in the first survey year and half in the second survey year. The remaining 186 areas, comprising less populated areas and nonmetropolitan areas, were divided into two 93-area groups, each of which was covered in 1 of the 2 survey years. From each of these 186 areas, a primary sampling unit (PSU) was randomly selected using a controlled selection procedure to insure proper geographic distribution.

Housing units within the 216 geographical areas were further assigned to housing unit strata. Occupied housing units were stratified by income level, housing tenure, and size of primary family. Vacant housing units were assigned to certain other strata as were institutional persons, i.e., that portion of the population living in rooming or boarding houses or in doctors' or nurses' quarters of general hospitals. The selection of housing units was independent from the selection of institutional persons. The actual sample of unclustered housing units was selected systematically by computer from the 1970 Census 20-percent sample data file which included those households completing the long form questionnaires. Approximately 30,000 housing units were designated for interview for each year of the 2-year diary survey. These units include the augmented sample of housing units which were to be visited during the 4-week period prior to the Christmas and New Year's holidays for each year. Additionally, they include a number of newly constructed housing units which were selected to update the sample for the period from the 1970 Census to the time of the interview for the diary survey. These new units were selected from reports of building permits issued for privately financed residential construction and were sampled independently within each PSU.

The sample of housing units for the Quarterly Interview Survey was also selected by computer from the 1970 Census 20-percent sample data file which included those households completing the long form questionnaires. Approximately 25,000 housing units were designated for interview over the 2-year quarterly interview survey collection

1

Major contributions to the design were made by Morris Hansen, Benjamin Tepping and Sid Jaffe of Westat Incorporated.

period with newly constructed housing units included to update the sample for the period from the 1970 Census to the time of the interview as for the Diary Survey.

4. Estimation of weights

Weights were assigned to each consumer unit (CU) participating in the survey and were products of several factors. One factor assigned was the inverse of the probability of selection of the housing unit, being equal to two times the initial sampling rate and reflecting the splitting of the sample into two 1-year components. For questionnaires which could not be collected from occupied sample households because of refusals or because no one was home, or from CU's within households where there was more than one CU, a complex noninterview adjustment was made to correct for these deficiencies. Additional factors included ratio estimation in non-self-representing PSU's for color and residence; ratio estimation in all geographic areas for age, sex, and color to known civilian noninstitutional population controls; and a CU adjustment based upon family composition.

B. CPIR PRICING AREA DESIGN

The old CPI was priced in 56 sample cities representing the urban population of the U.S.A. However for the CPIR the number of pricing areas was increased to 85 PSU's. The construction of this design is summarized as follows: The entire country was divided into Primary Sampling Units (PSU's). Each SMSA was a PSU. The remaining counties were grouped into contiguous counties. The PSU's were stratified into 85 strata from which one PSU per stratum was selected.

The stratification was done by combining similar PSU's according to the following characteristics:

- a. percent population increase 1960 to 1970,
- b. major industry, c. percent nonwhite and
- d. percent urban.

Major importance in determining the stratification of PSU's was given to percent population increase (or decrease). This design resulted in 27 strata with one pricing area per stratum (self-representing PSU's) and 58 non-self-representing strata. In selecting the one sample PSU for each non-self-representing stratum, a controlled selection program was used to insure the sample areas were properly distributed geographically across States. The 58 area design contained at least four pricing areas within each of the 12 region-city-size-classes. The number of areas was increased to adequately support the region-city-size-classes publication requirements and to reduce the between-PSU component of variance for the CPIR index.

C. POINT-OF-PURCHASE SURVEY (POPS)

In the spring-summer of 1974 a second household survey the Point-of-Purchase Survey was conducted by the Census Bureau to provide the sampling frame of outlets for food and most commodities and services to be priced in the CPI and to provide demographic data for classification of the households reporting an expenditure for an outlet. The Point-of-Purchase Survey was conducted in the 85 PSU's defined for the CPIR design. The commodities and services for which sampling frames were developed in each PSU included Food,

Apparel, Drugs, Personal Care items, Household furnishings and Housekeeping Supplies, Beverages, most Medical Services, Sports Equipment and Automobile components. Expenditures and name and location of the place of purchase were collected for approximately 100 relatively broad categories of expenditures from one week to two years depending on the expected frequency of reporting. There were two groups of categories, one set given to $\frac{1}{4}$ of the sample households and the second set given to $\frac{3}{4}$ of the sample households. The combination of sample size and reference period for a given POPS category was designed to generate approximately 6 or 12 gross outlets reported for a given PSU/POPS category.

For POPS the national sample size was 23,000 designated housing units. Since separate frames of outlets were required for individual CPIR pricing areas (PSU's), the sample size is not self-weighting across PSU's, but within a PSU, the households are selected with a uniform probability. In '60 of the 85 areas a highly clustered sample of households was selected. The assumption was that if families tended to buy in the areas where they live, the outlets given as responses to the survey would also be clustered. In order to increase the expected chance of clustering outlets responses, the household clusters were formed around known shopping complexes. Within these large clusters are called SSU's, a sample of ED's was selected and within the selected ED's (Census enumeration districts), the sampled households were dispersed evenly. Five housing units were selected in each ED and since the total sample size per cluster was desired at 40 housing units, about eight ED's were in sample in each cluster.

In the initial formation of the SSU's, all known shopping complexes were spotted on tract maps. These shopping complexes were the Central Business Districts, Major Retail Centers, (as defined by the 1967 Census of Business) and other shopping centers as defined by the 1970 Directory of Shopping Centers. Census tracts were then grouped into SSU's around each of these spotted shopping complexes. The guidelines for forming these SSU's were as follows:

1. All tracts containing part of the shopping complex should be in the same SSU.
2. Each SSU should ideally have only one shopping complex.
3. The SSU should have a minimum of 2,500 housing units.
4. The SSU should be more than two miles square.

This operation did not account for all the tracts in a PSU. The remaining tracts were merely put into contiguous groups according to above guidelines 3. and 4. as well as the following:

5. If possible all tracts in an incorporated or unincorporated place outside the central city of the SMSA and greater than 2,500 population should be an SSU.
6. SSU's should not cross the central city or urbanized boundaries. The Census Bureau sampled new construction units in each PSU in their normal manner (off permits registers) without regard to SSU.

For the remaining 25 smallest PSU's, the definition of SSU's was skipped and sampling of ED's

was the first stage of selection.

D. NON POINT-OF-PURCHASE - GENERAL

For approximately 40% of the CPIR expenditure weight, the Point-of-Purchase Survey did not provide universe of outlets. This section of the CPIR is called Non-POPS and encompasses significant sections such as rent, property tax, mortgage interest, house prices, utilities, transportation, insurance and several miscellaneous categories. For each of these components, unique universes were constructed or purchased for sampling or various imputation techniques were used.

The Non-POPS frames can be divided into two major categories; those dealing with household surveys and those derived from other sources which we will call Non-household Non-POPS. The data sources, outlet sampling units, definition of measure of sizes and pricing units collected are available on request from BLS.

III. CPI SAMPLE DESIGN

A. DEFINITION OF SAMPLE DESIGN PROBLEM

The initial plans of BLS were to expand the present CPI population coverage (urban wage and clerical worker consumer units) to include all civilian noninstitutional consumer units (about 80% of total the population). The specifications for the POPS survey were predicated on this assumption. Subsequently, it was decided to continue publication of an index for the urban wage and clerical worker population (about 40% of the total population) as well as to publish the more extensive index, both with the same degree of reliability as the current index. Hence, the sample design problem was to select two samples of items and outlets such that two indexes had at least the same reliability at the national level as the present CPI and also meet cost constraints which except for computer funds were to be held constant.

The following describes the sampling procedures to select the items, outlets, and specific store items within outlets, for the two family indexes.

B. STRUCTURE OF COSTS WEIGHTS AND SELECTION OF ELI's

The relationship between the item sampling frames (results of Quarterly and Diary Surveys) and the outlet sampling frames (results of POPS or Non-POPS) is as follows: The expenditure data for the CPI was constructed from the Diary and Quarterly Consumer Expenditure Surveys. Only the first year Diary or Quarterly data was used for selection of items, since this data was all that BLS could process prior to item selection. The Diary data was the sampling frame for items frequently purchased, such as food and personal care items. The Quarterly data was used for all other selections.

The basic structure created for both the Diary and Quarterly was to define 68 Expenditure Classes (EC's) which are BLS's primary definition of publication levels of indexes. Within each EC the expenditures defined therein were grouped into one or more item strata (265). Within each item stratum, one or more substrata called Entry Level Items (ELI's) were defined (382). The ELI's are the ultimate sampling unit selected in Washington and are used by the data collectors as their initial level of item definition within an

outlet. These ELI's are relatively broadly defined groupings of items and allow the possibility of pricing many different kinds of specifications or specific store items rather than a single specification as previously used in the CPI. The ELI's are mutually exclusive and account for all consumer expenditures reported in the Expenditures Surveys.

This structure of the Diary and Quarterly into item strata and ELI's is the same for all regions and Market Baskets. However, for sampling purposes BLS tabulated four regional universe market baskets for each population to reflect regional differences rather than the national market basket approach as used in the old CPI. Within each region, BLS made eight independent selections of ELI's within each item stratum forming eight "half-samples" of ELI's. These half-samples were distributed among the CPI pricing areas (PSU's) for pricing within the region in such a manner that the variance estimation procedure (replication) would measure all components of variance of the index.

The reason 32 half-samples were selected for each population was to reduce significantly the correlation of pricing the same items between pricing areas. The optimum strategy would have been to make the number of selections equal to the number of half-samples in the region but workload requirements dictated that a compromise be made; however, the 32 selections provide the major portion of the possible reduction of the correlation between item samples.

Because of the requirement of publishing two family indexes, the following technique was used for item selection. Each selection of ELI's within an item stratum was made initially for the urban wage and clerical population (W) proportional to the relative expenditure of ELI's within an item stratum for the W population. Then using a technique developed by Nathan Keyfitz to maximize the overlap of ELI's between populations and maintain the correct probabilities of selection, a second selection of ELI's for the all urban consumer unit population (U) was made proportional to the relative expenditures of the U population.

Each ELI was defined to be in one and only one POPS or Non-POPS category so that the integration of the ELI sample and the outlet sample was by the ELI/POPS or Non-POPS category concordance. For a given population/half-sample, a single selection of outlets in an outlet category was used to price up to seven ELI's for Commodities and Services (C&S) items and 14 food ELI's. Thus, the selection of the ELI identified which outlet category was to be used for outlet selection for the PSU/half-sample/population.

IV. OUTLET SELECTION

A. POPS CATEGORIES

The following approach was used for outlet selection for frames developed from the POPS Survey. For a given PSU/category the total expenditures for a given outlet were edited and insured a minimum chance for selection or were restricted to 20 percent of the total expenditure for the U population of the PSU/category. A systematic selection of outlets reported for a given POPS category for the W population was

made where the measure of size for each outlet was proportional to the average daily expenditure reported for the outlet by all households of the W population. The outlets for the U population were then selected by using a Key-fitting technique to recompute the measure of size for every outlet in the universe and the sample outlets for the U population were then selected by a repeat of the systematic selection using the new measures of size. The specific algorithm and its justification is available from BLS and has been prepared by Harry Marks. As a sidelite, the outlet selection procedure permitted multiple selections of the same outlet if the proportion of expenditure was great enough.

B. NON-POPS NON-HOUSEHOLD SAMPLING FRAMES

The following approach for outlet selection for Non-POPS Non-Household categories was used for outlet selection. The outlets were ordered by whatever measure of size was available from the frame and a systematic sample of outlets was selected proportional to the measure of size for each PSU/half-sample. No demographic information existed on any Non-POPS frame so that the same sample of outlets and items priced is used for each CPIR population.

C. INTEGRATION, CREATION OF QUOTES AND SAMPLE SIZE

When the item strata were defined, an initial allocation of quotes per outlet hit was made proportional to the relative percent of expenditure of the item strata at the national level for the U population with the constraint that each item strata had at least one quote sample per half-sample/population. Minor adjustments were then made to reflect regional differences, no adjustments were made for population differences of percent of expenditure since the U and W distributions are relatively similar.

A quote sample is defined usually for a given PSU/half-sample/population as one quote per outlet hit for a specified number of outlets per outlet category. The specified number of outlet hits desired was 2 or 3 and 4 or 5 for commodity and service categories and food categories respectively. Each of these desired sample sizes were increased at the time of sample selection by 1 or 2 outlet hits to allow for nonresponse for any reason.

For food for both populations combined, about 58,000 prices are collected each month in about 2,300 outlets. For other commodities and services 675,000 prices per year are collected in 22,300 outlets. Some of these prices are collected monthly and the remainder are collected bimonthly.

The samples of ELI's and outlets were merged and listings of the sample outlets and the sample ELI's with the assigned number of quotes designated within the outlet (determined by the product of the number of quotes/outlet hit for item strata and the number of times the outlet was selected for POPS category) were prepared on a "facesheet" for the data collector.

D. WITHIN OUTLET SELECTION FOR SPECIFIC ITEMS

For each ELI whether it was from a POPS or Non-POPS category, the selection of a specific store item by a data collector was done using multi-stage probability selection techniques with mea-

asures of size proportional to percentages of dollar sales usually provided by the respondent of the outlet. This procedure is new for the CPI which formerly asked for the "volume seller" for a national tight specification describing a specified quality of item required for pricing.

For example, the old procedure required an outlet be selected to price a shirt of a specified grade of fabric (a tight specification as opposed to a broad ELI-shirt). Then, if more than one shirt met the required specifications, the "volume seller" brand and style was selected for pricing. In the revision, an outlet is selected for shirts. A specific shirt is chosen by sampling proportional to the percentage of sales of all the categories and types of shirts in the outlet. The detailed description of the shirt is recorded for future pricing.

To perform this operation, the data collector is provided with a checklist that includes all the price determining characteristics of items defined within the ELI. In addition, the data collector is given the definition of the ELI, suggested stages of groupings of items to aid in quickly selecting a specific store item and a series of worksheets on which to define the categories of items, post the probabilities and identify the next category within which to select the specific store item by use of the random number table on the worksheet.

In developing this procedure, it became necessary to provide the data collector with several alternatives for defining the categories and obtaining the percentage of dollar sales or approximations to those sales. The procedures developed to obtain the proportion of sales were: 1. Obtaining the proportions directly from a respondent. 2. Ranking the categories (by respondent) and then obtaining the proportions directly or using preassigned proportions. 3. Using shelf space to estimate the proportions where applicable. 4. Using equal probability if all else fails. 5. To define the categories, direct responses from the respondent as to what he sells or an inventory technique.

The procedures make possible an objective probability sampling of quotes throughout the CPIR. They also allow broad definitions of ELI's so that the same tight specification need not be priced everywhere. The wide variety of specific items greatly reduces the within EC item variance and allows a substantial reduction in the number of quotes required to obtain the same reliability as the old index. A second important benefit from the broader ELI's, along with the POPS categories, is a significantly higher probability of finding a store item within the definition of the ELI within the sample outlet.

V. ESTIMATOR OF $R_{t,t-1}$

With the use of the POPS survey and of Non-POPS Frames and probability sampling for selecting outlets and quotes in the CPIR, the determination of the estimator of the price relative $R_{hzt,t-1}$ is a function of the measure of size used to select the sample of outlets and quotes. Two types of measures of size were used in the CPIR:

1) Expenditures ($P_0 Q_0$) for all food and commodity and service items, and 2) Equal measures of size for the rent/property tax samples (i.e., $Q_0 = 1$).

For ELI's for which the outlets and quotes were selected by probability sampling where the measure of size for selection was expenditures, revenues, or any substitute for which the assumption of proportionality to revenue is acceptable, the estimation of the price relative of change over a period can be estimated by a ratio of two long term relatives.

$$R_{zt,t-1} = \frac{R_{zt,0}}{R_{z,t-1,0}}$$

For each item stratum, a relative $R_{zt,t-1}$ is computed regardless of the number of eligible quotes. Following the computation of the relatives, the number of quotes used in the computation are examined. If the number of quotes is less than the minimum number, an average of relatives from "adjacent" item strata are used to estimate the change. Detailed specs are available on request.

A. RELATIVE COMPUTATION

Each pricing period (monthly or bimonthly) for a given item stratum/population, relatives are computed. For computation of the relatives only the quotes obtained during the current pricing period and the previous pricing period are used which existed and are comparable both periods. The same internal weights are used for both the current price and the previous price when computing the relative. Any quote for comparable specifications which has a current price and a previous month's price is considered useable. The previous month's price may be an actual price, an imputed price or a quality adjusted price.

The estimate of a one price period change (t-1 to t) for a given stratum for the Rth replicate is the ratio of the following two long-term relatives:

$$R_z R_{t,t-1} = \frac{R_{z,t,0}}{R_{z,t-1,0}} = \frac{\sum_{h=1}^n \sum_{i=1}^n W_{hijRt} (P_{zijt}/P_{zijo})}{\sum_{h=1}^n \sum_{i=1}^n W_{hijRt}} \bigg/ \frac{\sum_{h=1}^n \sum_{i=1}^n W_{hijRt} (P_{zijt-1}/P_{zijo})}{\sum_{h=1}^n \sum_{i=1}^n W_{hijRt}}$$

where: The subscript h indicates the possibility of summing across different PSU's in B, C and D areas;

P_{zijt} = is the price for the ith quote obtained in the current month t in the jth outlet for the zth item stratum;

P_{zijt-1} = is the price for the ith quote obtained in the previous pricing period (t-1) in the jth outlet for the zth item stratum; or an imputed or quality adjusted price as defined in paragraph V.C.2;

P_{zijo} = is the base period price for both quotes;

W_{hijRt} = is the current price period weight for the quote. This weight is defined as:

the product of a basic weight, a duplication control, a value for the percent the ELI is of the POPS category, and the number of useable quotes in time t for the item stratum. The final weight for a quote, which is computed each month, is a

function of all of these factors for the appropriate population.

1) Basic Weight (W)

Within the kth POPS category/population for each PSU, an estimate of the total daily expenditure for the POPS category/population.

2) Duplication Control (f_{1j})

This factor is the adjustment needed for any field or special subsampling of outlets or quotes for the POPS category/population and ELI.

3) Percent ELI is of POPS Category (α_{ij})

For each ELI/outlet where the ELI is defined as a proper subset of the POPS Category, an estimate of the proportion the ELI is of the POPS category is usually obtained by the data collector. The values will range from 0.0 to 100.0.

There is no distinction between populations for the α_{ij} 's.

4) Useable Number of Quotes (N_{hilt})

This is the number of quotes which are determined as useable in time t for a given ELI/half-sample/population. By using this variable, a form of noninterview adjustment is made for the quotes which are not obtained each month which should have been priced, and to some extent compensates for this bias.

5) Final Weight

The final weight for each quote a given population each month is the product of these elements:

$$W_{hijt} = W_{kl} \times f_{1j} \times \alpha_{ij} \times \frac{1}{N_{hilt}}$$

All weights are maintained separately for both populations to provide flexibility in the system.

Computation of Effective Prices:

For food some prices are reduced to what are called "effective" prices. These are computed by dividing the price obtained by the product of the quantity of units priced and the size (in ounces) of one unit.

Base Period Prices for Relative Computation:

The base period prices used for each quote are also in terms of effective prices. They are the actual effective prices collected in the link month, which is the first month that relative computation is "real" (Jan. to June 1979) and used for moving the cost weights.

B. IMPUTATION FOR REPRICING

Following the computation of the one pricing period relative R (t, t-1) for each item stratum or collapsed set of item strata, the prices for quotes which were not obtained in period t are imputed based on the price reported in t-1 and the relative R (t, t-1) so that a price for the quote will be available for comparison in the next pricing period (t+1).

The imputed price $P_t^* = R_{zh}(t,t-1) \times P(t-1)$. The relative to use for imputation is the relative used to move the cost weight for the item stratum.

C. LINKING

1. No Overlap Price

For quotes which are not comparable because there is a change in the specification (substitution) being priced between period t-1 and t and no

quality adjustment or overlap price is obtained, the process of linking is used to prepare the quote for future use in measuring the index. The new specification price is not used in the current estimate of the relative. To execute the link, an estimate of the long term change for the previous pricing period ($R_{z,t-1,o}$) for the item stratum and the "current" one month pricing period relative ($R_{z,t,t-1}$) are required. The value of ($R_{z,t,t-1}$) is the current estimate. Linking is the estimation of a new base period price ($P^*_{zhi,o}$) for the new specification as follows:

$$P^*_{zhi,o} = P_{zit} / (R_{zt-1,o} \times R_{zt,t-1}) \quad (A)$$

The $R_{zt-1,o}$ is estimated by using the ratio of cost population weights for the item stratum/market basket/cycle where

$$R_{zt-1,o} = \frac{CW_{z,t-1}}{CW_{z,o}}$$

where $CW_{z,t-1}$ is the cost population weight of the previous pricing period for the item stratum/market basket/population/cycle. $CW_{z,o}$ is the cost population weight of the previous pricing period when the item stratum/market basket/cycle is linked to the old index. This is not the base period cost population weight.

2. With Overlapping Price

When a substitution occurs or the quote has been coded by the commodity specialist as "not comparable, link" and a price is obtained for both the old and new specification (overlap pricing) the estimation of the new base period price (link) is done at the quote level to maintain the best estimate of continuity for the quote. The new base period price (P^*_{io}) is estimated by multiplying the new price (P^*_{it}) by the inverse of the old price (P_{it}) to its base period price (P_{io}), that is,

$$P^*_{io} = P^*_{it} (P_{io}/P_{it})$$

The linking of quotes with overlap prices is done before relatives are computed.

3. Mechanics of Quality Adjustment

Frequently, perhaps each pricing period, some quotes will have had substitutions (new specifications defined by the field) for which the commodity specialist is able to adjust the old specifications' price for the change in "quality" that gave rise to the new specification. The resulting adjustment factors are either additive or multiplicative. Prior to the computation of a relative, a code is available to the commodity specialist with which to indicate whether they are applying an additive or multiplicative factor. Adjustments are made to the previous pricing period's price and stored separately.

VI. HOUSEHOLD SURVEYS IN THE CPIR

A. RENT AND PROPERTY TAX - DIRECTLY COLLECTED DATA

A sample of housing units was selected in 85 CPIR PSU's from which a subsample renter units is used to measure rent change and another subsample of owner units is used to measure change in property tax. All housing units in each sample (rent and property tax) are used for both population indexes.

1. CPIR Rent Sample Design

The CPIR Rent Sample is a sample of 21,000 desig-

nated housing units. These housing units were divided up into six panels, one panel being interviewed every month and a given panel being interviewed every six months. At the time of data collection the current and previous month's rent is collected. This is to provide a measurement of one month change as well as a six month change.

Taking a weighted average of these relatives defines a composite estimator. This composite estimator enables the measurement of rent to be an up-to-date measure of rent change, not lagged estimate as when using only a six month estimate of change, as in the old CPI. In addition it is cost effective since fewer households need to be interviewed to obtain the same reliability. The same rent relative is used for both populations' indexes in the CPIR.

In general,

$$I_{ht} = WB_{ht-1} r_{ht-1} + (1-W) B_{ht-6} r_{ht-6}$$

where the cost population weights for the previous month are (B_{ht-1}) and the previous six months (B_{ht-6}). The final one month estimate of rent price change $R_{t,t-1}$ is estimated by taking the ratio of two composite estimates for the stratum.

$$R_{t,t-1} = \frac{I_{ht}}{I_{ht-1}}$$

2. Property Tax Sample Design

From the sample of screening schedules identified as owner units or tenure "not available", a sample of owner units built before 1970 were selected so that an average of about 2.5 owner housing units were selected per segment approximately 17,000 designated housing units.

D. MORTGAGE INTEREST AND HOUSE PRICES - INDIRECTLY COLLECTED HOUSEHOLD DATA

For Mortgage Interest and House Prices, data is not derived directly from a sample of households but from secondary sources.

1. Mortgage Interest

a. Sample Designs

The Federal Home Loan Bank Board (FHLBB) provides to BLS a sample of mortgages each month from a national sample of about 1,600 mutual service and commercial banks, savings and loan associations, and mortgage companies. Three lender size strata of the different banks were defined with a differential sampling rate used in each across the nation. All transactions which occur the first week of the month within the sample units are in sample. For the FHA/VA component of mortgage interest BLS enters the FHA/VA mortgage interest ceiling for these interest rates.

b. Estimation

For each pricing area for each of the lender strata an average interest rate is computed for the month and a weighted average of the average interest rates is computed for month t and t-1 to compute the estimate of price change for mortgage interest cost for a fixed reference period (1972-73). The final relative is the product of the change in mortgage interest rate times the change in house prices.

2. House Price

Further information on House Price can be obtained from the Bureau of Labor Statistics.