Revision of the CPI housing sample and estimators

Changes in concepts and methods being readied for the 1998 revision should smooth the process of index production for shelter services

Frank Ptacek
and
Robert M. Baskin

The Consumer Price Index (CPI) currently consists of seven major categories, with the large aggregate grouping called 'housing' representing 41 percent of the total index. Within the 'housing' category, most of the relative importance belongs to the index for 'owners' equivalent rent', also known as 'rental equivalence', and that for 'residential rent,' for which data are obtained through the Housing survey. These two items represent 20 percent and 6 percent, respectively, of the total CPI. The remaining 15 percent accounted for by the 'housing' group consists of many indexes, which are handled through the Commodities and Services survey. (For more detail see "Changing the Item Structure of the Consumer Price Index," pages 18–25.)

In addition to data on residential rents used to calculate changes in rents for the 'rent' index, the Housing survey also collects data for owned homes for use, in conjunction with the rent data, in calculating of changes in the rental value of owned homes for the 'rental equivalence' measure. Clearly, the rental value of owned homes is not an easily determined dollar amount, and Housing survey analysts have spent considerable time and effort in estimating this value. As a result, determining 'rental equivalence' is an important issue in the upcoming 1998 revision of the CPI.

The revision has provided a window of opportunity to initiate an extensive redesign of the Housing survey. Beginning in 1999, the CPI for 'rent' and for 'rental equivalence' will be based on a new sample design and estimation methodology. Major technological advances and improvements in the operational processes also will

be implemented in the survey. More specifically, these changes include:

- improved design for stratification and selection of housing units, based on the 1990 Decennial Census;
- improved methodology for estimating changes for the 'rent' and 'rental equivalence' indexes; and
- technological innovations and modernization of all processing systems used for the production of both indexes, including the new Housing Computer-Assisted Data Collection System.

This article describes the systems, data bases, and procedures that are being developed for the upcoming revision of the Housing survey, and explains the advantages to be derived from each.

Background

The 'rental equivalence' approach to measuring price change for owner-occupied housing was implemented for the CPI for All Urban Consumers (CPI-U) in January 1983, and for the CPI for Urban Wage Earners and Clerical Workers (CPI-W) in January 1985. In essence, 'rental equivalence' measures the change in the amount a homeowner would pay to rent, or would earn from renting, his or her home in a competitive market. It is a measure of the change in the price of the shelter service provided by owner-occupied housing. When initially introduced, the 'rental equivalence' index was moved (that is, changes were applied) by reweighting the rent sample to represent owner-occupied units. The preferred methodology would have been to

Frank Ptacek is chief of the Housing Section, Division of Consumer Prices and Price Indexes, and Robert M. Baskin is a mathematical statistician, Division of Price Statistical Methods, both at the Bureau of Labor Statistics.

match owner units to renter units and use those more specific rent changes to calculate changes in the rental value of owner units. The reweighting approach was taken because an owner sample could not be selected and available for use before the CPI was last revised in 1987.

Since January 1987 (the 1987 revision), the 'rental equivalence' index movement has been based on changes in the implicit rent of owner units. These implicit rents are moved by the changes in the pure rents (which exclude the cost of any utilities included in the rent contract) of matched rental units. The implicit rents are estimated by the owners in the CPI owner sample, and those implicit rents are then moved by the specific rent changes for renter units with similar characteristics (owner/renter matching). The characteristics include location, structure type, and other general traits such as age, number of rooms, and type of air conditioning.

The 'rent' index measures the changes in rents, specifically "contract rents," paid by tenants or received by landlords. "Contract rents" are the payments for all services the landlord provides in exchange for the rent. For example, if the landlord provides electricity, it is considered part of the contract rent. The data collected for the 'rent' index consists of rent, rent reductions, extra charges, and information concerning the utilities, facilities, and services received for the rent.

The 'rental equivalence' estimator

In any properly designed statistical study, samples are selected to support the estimation process that is planned. The 1987 revision Housing sample was selected to support the estimation of the 'rental equivalence' index through the use of implicit rents for owner-occupied units and the movement of the implicit rents through owner/renter matching.

For purposes of the upcoming 1998 revision, the decision was made to drop the owner sample and return to the methodology that was used for the 'rental equivalence' index when it was first introduced—that is, the reweighting of the rent sample to represent owner-occupied units. This decision was made for several reasons:

• In order to move the implicit rents of the owner sample, it is necessary to find renter-occupied units in mostly owner-occupied neighborhoods. Locating such renter-occupied units has proved to be a time-consuming and sometimes futile task. Also, the loss of renter-occupied units in these neighborhoods, through conversion to owner occupancy or demolition, can have a large impact on the sample. In 1992–1994, CPI staff endeavored to locate and augment the sample with additional renter-occupied units in the mostly owner-occupied neighborhoods but met with only partial success. The small sample size of renter-occupied units also contributed to increased sampling variance.

• Even though different sampling methodologies could be employed, there was no certainty that sufficient numbers of renters in mostly owner-occupied neighborhoods could be found to support owner/renter matching. It was also clear that keeping owner/renter matching, but relaxing the matching criteria to increase the number of matches, would reduce the desired specificity of the matches and, as a result, the desired specificity of the rent movements.

• Moving implicit rent estimates by matching renter and owner observations is inherently a reweighting of the rent sample. Given that the samples for the 1998 revision are being selected to support the reweighting methodology, CPI staff was able to derive the owner weights from 1990 census data. These weights are clearly much better than those derived when 'rental equivalency' was initially introduced.

• A large portion of the 1987 sample is devoted to owners, to support the estimation of initial implicit rent. By dropping the owner sample, the field staff will not have to initiate, price, and maintain an owner sample. Again, because the 1998 revision samples are being selected to support the reweighting methodology, mostly owner-occupied neighborhoods, while assured of having the correct probability of selection, do not have to be over-represented, as they are in the 1987 sample. The cost savings in not having an owner sample will be spent on improving the renter sample.

• Because owner/renter matching, a very complicated process, will not be needed in the revision 'rental equivalence' estimator, calculation of the index has been greatly simplified. (See the section on estimation below.) In addition, the entire Housing computer system has been simplified because data no longer have to be stored and processed for two types of housing units (renters and owners) with different characteristics. The streamlined data processing system will be more easily maintained, and will be easier to change as the need arises.

Sample selection

Geographic stratification. Research performed by BLS using 1980 and 1990 census data indicates that geographic location is the most important variable (that is, it accounts for most of the variance) in determining rent change. Once geography is taken into account, only rent level is significant in predicting rent change. The percent of owner-occupied units in a neighborhood, which was a key stratification variable in the 1987 sample selection process, proved to be of little importance in explaining change.

Geographic software, which was not available for the 1987 revision, allowed stratification by geography for the 1998 revision. The geographic stratification accomplished five goals:

1. It helps ensure sample coverage for the major charac-

teristics (geography and rent level) that are correlated with rent change.

2. It is felt to be the best way to correlate renter-occupied units with owner-occupied units in the same neighborhood, in order to produce the 'rental equivalence' index.

3. Housing units constructed after 1990 can be located and assigned to the existing geographic strata, as described below.

4. Because goals 1 through 3 will be met, there should be a reduction in the sampling variance of the 'rent' and 'rental equivalence' indexes.

5. It sets up a stratification structure that will allow the rotation of Housing samples on a rolling basis, thereby distributing the introduction of future census samples over an extended period.

The Housing sample for the 1998 revision is a stratified cluster sample, which represents housing units built before 1990. Housing units built after 1989 are handled through the New Construction survey, as described later in this article. Using data from the 1990 Census of Population and Housing, CPI analysts divided the Primary Sampling Units[1] (PSUs) into geographic neighborhoods (segments). The segments are small contiguous groups of census blocks (sectors). The segments contain at least 50 housing units in the larger PSUs and at least 30 units in the smaller PSUs. These segments are stratified by location within the PSU. Six geographic strata were formed in each PSU. Once geography is taken into account, only rent level is significant in predicting rent change, so the stratification boundaries were determined using information about population and median rent level.

In the first step of the stratification process, a box is found in the geographic center of the PSU, so that about one-third of the population is contained inside the box. The box is then split into two strata. Whether the split is by latitude or longitude is determined by rent level. The split that maximizes the difference in median rent level determines strata 1 and 2. Then, the four noncentral strata are determined iteratively in a similar fashion. The entire noncentral part of the PSU is split into two parts, either by latitude or longitude. Once the first noncentral split is determined, a split perpendicular to the first split is made within each half.

Exhibit 1 shows the six geographic strata in the St. Louis, MO-IL PSU. According to this map, the central box was split by longitude, and then the entire noncentral part was split by longitude, with each half then split by latitude. While rent level, as well as population, was used to determine the geographic strata boundaries, the resulting strata are purely geographic divisions of the PSU. Two of the strata correspond roughly to the most densely populated part of the PSU, and the other four strata correspond to surrounding suburban areas.[2]

Exhibit 1. **Segment sample selection for St. Louis, MO–IL**
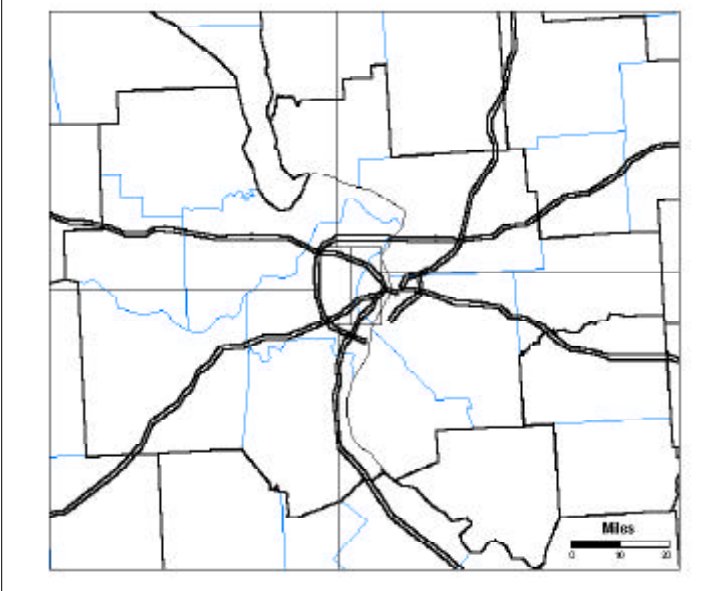
Weighting during segment sample selection. CPI analysts then selected segments in the strata to represent housing units constructed before 1990. In the 1987 revision, segments had been selected with probability proportional to size, the size measure being the number of housing units in the segment. When the number of units is used as the size measure, smaller, less expensive housing units (in apartment complexes, for example) have the same probability of selection as more expensive single-family units. Because the 'rent' and 'rental equivalence' indexes are measures of the change in the price of the shelter service provided by renter-occupied and owner-occupied housing, it was felt that higher expenditures (rent levels) should have a higher probability of selection. In the 1998 revision, therefore, segments were again selected with probability proportional to size, but the size measure was estimated *expenditures*.

In the segment selection process, the segments are ordered within each stratum by county and then by segment rent level within county. Because the segment selection is systematic, this guarantees that not all high-rent or low-rent segments are chosen.

Each segment has a probability of selection within the stratum (P_s) that is the ratio of the cost of housing in the segment relative to the cost of housing in the stratum. Therefore,

$$P_s = TC_s / \sum_{s \in S} TC_s$$

where S = stratum; and

TC_s is defined below.

Each segment also has a weight (W_s), which is the reciprocal of the probability of selection. Therefore,

$$W_s = \sum_{s \in S} TC_s \, / \, TC_s$$

The cost of housing in the segment is the cost of rented housing in the segment (RC_s) plus the cost of owned housing in the segment (OC_s). The RC_s is the number of rented housing units in the segment (R_s) times the average rent value within the segment (RR_s). The OC_s is the number of owned housing units in the segment (O_s) times an estimated average owner equivalent rent value within the segment (IR_s). This gives segments with higher-valued units (that is, higher rent levels) a higher probability of selection and a lower segment weight. The relationship among these variables is given by:

$$TC_s = RC_s + OC_s = R_s * RR_s + O_s * IR_s$$

The estimated average owner equivalent rent value (IR_s) was determined by a nonlinear regression of the 1990 census owner value within census blocks on the 1990 census average rent value within the same census block:

$$y = b_0 * (1\text{-}\exp(\text{-}b_1 * x)) + e$$

where y = average rent;

x = average owner value; and

z = average implicit rent.

The actual regression coefficients (b_0 and b_1) were determined uniquely within each PSU.

Because rents are not volatile, the Housing sample is divided into panels; one panel is priced each month and each panel is priced twice a year. For example, panel 1 is priced in January and July, panel 2 in February and August, and so on through panel 6. The segments within the strata are assigned to these panels. These assignments are made such that each panel has a representative subsample of the PSU. Because each panel is representative of the entire sample and there is never an off-cycle month for the Housing survey, a panel of data provides sufficient information for monthly publication of the 'rent' and 'rental equivalence' indexes. Primary segments were selected within the PSUs in multiples of 36, so that each combination of stratum and panel had the same sample size.

Other segment sample selection outputs.   About 10,000 segments have been selected in the PSUs and the 1998 revision Housing unit sample is designed to consist of approximately 50,000 rental units. CPI analysts have computed sampling rates for each segment, so that the sample design will be realized after the listing, sampling, and screening processes, as described below, are completed. These sampling rates will be used during the listing process to select the addresses that will be screened for use in the Housing sample.

Segment-level information from the selected segments will be provided to the mapping system. This information allows the production of all maps required by field staff to locate the segments within the PSUs.

New construction augmentation.   The augmentation of the Housing sample with newly constructed housing units is not part of the segment sample selection process, but it is discussed here because these housing units will fit neatly into the geographic stratification of the Housing sample. The Census Bureau will supply to BLS a sample of address records from building permits, representing housing units built after 1989. (BLS calls this list of address records the New Construction sample.) BLS expects to receive about 1,000 address records per year from the Census Bureau, with 20 percent of these yielding usable renter-occupied units after they have gone through the screening process. Once they receive the new construction sample, CPI analysts will assign each address record to one of the six geographic strata based on the zip code. They then will allocate the new construction sample among the segments, using the census sample design and zip code.

Other system modernizations

Mapping system.   Sets of maps are needed to help field staff locate the sectors within the segments that must be recorded in the listing process described below. In previous CPI revisions, maps with PSU, segment, and sector identifiers, along with street names and boundary information, were produced by hand in Washington and provided to the field staff. Because corrections to the maps were entered and kept by the field staff, the BLS Washington Office did not maintain an updated set of maps for all PSUs, segments, and sectors. For the 1998 revision, the CPI systems staff has developed a system to produce sets of maps, using the Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER) data and commercial, "off-the-shelf" Geographic Information System (GIS) software. The process of segment sample selection and use of the Sample Maintenance and Control System (SMCS) described below will provide the information necessary to accurately specify the selected segments within each PSU/stratum and each sector within those segments. The mapping system will yield accurate, reproducible sets of maps with all necessary information for the field staff and will extract the defined limits (boundary information) for each sector from the TIGER data. These boundary data will be provided to the SMCS portion (see below) of the Housing data base for use in the listing process. The sets of maps will include:

- *Sector maps for each sector in the segment*—The maps will highlight the sector and denote the street boundaries of the sector, as well as local street names and/or highways surrounding the sector.
- *Segment maps for each segment in the PSU*—The maps will highlight the segment, as well as denote the local street names and/or highways surrounding the segment.
- *County maps for each county in the PSU*—The maps will highlight and label all selected segments in the county, as well as major highways, water, railroads, and place names within the county.

- Corrections and additions will be entered on the maps by the field staff, usually during the listing process, and a copy of the corrected map will be sent to the Washington Office, where the changes will be entered in the mapping data base by cartographers. New maps will be supplied to the field staff before pricing begins. This process will allow the Washington Office to produce updated maps upon request for all PSUs, segments, and sectors.

Sample Maintenance and Control System (SMCS). Previously, much of the sample information for the CPI Housing survey was maintained separately from the rest of the Housing data base, a situation that complicated sample administration. It also was hard to analyze the current sample using the Housing data base due to its panel structure. With the 1998 revision, however, sample information will become an important part of the Housing data base. The SMCS is a new, consolidated system that will be used to control the Housing Sample. It has five major functions:

- Storing and processing sample-related information;
- Providing sample data and information for review and evaluation;
- Controlling the timing of the activation of housing units for index use and the removal of units from index calculation;
- Measuring and monitoring the sample size at different stages in the Housing survey processes and keeping records of the sufficiency of results; and
- Providing standardized reports to the Washington Office and the field concerning the status of the samples and their stages of processing.

The SMCS performs these five functions during the four major stages of the sample's existence—as described below:

1) *Listing.* Listing is the process of recording the addresses of all housing units in a segment. The Housing data base will contain all segments within the PSUs. (These are the outputs of the sample selection process described above.) The SMCS will set numerous scheduling variables that will govern when the segments will be sent to the field for listing. Once a segment has been listed, a sampling algorithm is applied to the listed addresses to determine which of them will be selected for the next stage of the process, scope determination, or screening.

2) *Scope determination (screening).* The Housing data base will contain all of the addresses that were listed for the segment, as well as an indication of which addresses were selected for screening. The field staff will then perform additional data collection for the selected addresses through observation and by asking questions of eligible respondents. This additional data collection will determine whether the selected addresses are eligible (in scope) for the Housing sample (scope status).[3]

The Housing data base will hold all screening data and the scope status for the selected addresses.

Selected addresses that fail screening for some permanent reason usually are never visited again. Those that fail screening for some temporary reason will be screened again after a specific waiting period. Selected addresses that pass the screening criteria are considered in scope for the Housing sample and are eligible to proceed to the next stage of the process, initiation.

3) *Initiation.* Initiation is the initial collection of rent data, which consists of the specific housing services that are associated with the unit and the rent that is paid. These data are the basis for all calculations of rent change that will occur during the life of the unit in the Housing sample. The Housing data base will hold all initiation data.

4) *Pricing.* After initiation, the housing unit is priced "on-panel" every 6 months, so that price changes can be reflected through the price relative calculation described later in this article. The Housing data base will hold all pricing data. Occasionally, situations will occur during pricing that will affect the unit's scope status and, on a scheduled but infrequent basis, additional questions are asked to ensure that a housing unit is still in scope for the Housing sample. If changes occur, the SMCS will store those results and the units will be subjected to the scope determination process described above, based on their new scope status.

Because initiation and pricing data do not govern whether a housing unit is in the sample, the SMCS does not look at those data. However, it does control numerous scheduling variables that govern when the unit can be used in the CPI computations, when it should be dropped from those computations, and when new samples should be fielded to augment or replace deficient or outdated samples.

Computer-assisted data collection. A key element of the 1998 revision is the conversion of all data collection and transmission to electronic systems.

1) *Listing/sampling instrument.* The listing of the sampled segments is a highly labor-intensive portion of the

Housing sample preparation activities. In previous revisions, the field staff were required to manually list all of the addresses in the segment on paper for data entry in Washington. This process required many shipments between the field and Washington. After the segments were listed, the field staff would laboriously apply sampling rates to the listed data in the field (or the sampling rates would be applied in Washington) in order to randomly select those addresses to be screened and possibly initiated for the Housing sample.

The use of portable computers for these activities should result in substantial savings, because computer-assisted data collection will consolidate listing and data capture into one operation. The listing instrument for the 1998 revision has been designed to make listing as easy as possible for the field staff. Street names are automatically entered when the staff member selects a street name from the sector boundary data. The first street number is entered, and then is automatically incremented (or decremented) as specified by the field staff. As the street numbers change, the street names are copied to each line as the staff member lists the sectors within the segment. When the staff member is ready to move to another street, he or she can select the next street name from the sector boundaries and proceed. In areas in which the street numbers are not clearly marked or do not exist, the staff member must enter a description of the property. A computerized phrase builder has been designed to reduce the amount of typing required of the field staff as they enter these descriptions.

After the field staff member determines, with the assistance of the computer, that the collected data conform to expected results, the computer applies the sampling algorithm to the listing data. The result of the sampling will be a set of addresses that have been selected for scope determination (screening). All of the listing data are electronically transmitted to the Housing data base in Washington.

If time permits, the field staff member may choose to screen/initiate ("screnitiate") the selected addresses. The listing instrument will generate "screnitiation" schedules using data collected during listing. These schedules will be passed to the collection instrument, so that the field staff may screnitiate these addresses while in the segment. If time does not permit, the screnitiation schedules will be electronically transmitted to the field staff from the Washington data base when a segment is next on panel.

2) *Collection instrument.* In previous revisions, if the field staff applied the sampling rates, the **screening** of the selected addresses involved the tedious transcription of identifiers and listing data to blank screening forms. (When sampling was done in Washington, computer-generated screening forms were supplied to the field staff.) The field staff then obtained answers to various (screening) questions (through observation and through direct questioning of eligible respon-

dents), which would determine whether an address was in scope for the Housing sample. If the housing unit was found to be out of scope for some reason that was not likely to change, the field staff assigned a "permanently out of scope" response code and the unit was never visited again. (An example of this would be units in public housing projects.) If the housing unit was found to be out of scope for some reason that might change, it received a "temporarily out of scope" response code and the unit was sent out for another screening and possible initiation after a specific recall period. (An example of this situation would be a unit that is not the primary residence for the current tenant, but may be for some future tenant.) If the screening was incomplete, the address was returned to the field in 6 months for another screnitiation attempt.

Inherent in the structured Housing questionnaires (screening, initiation, and pricing) are various flow determinations (skip patterns), such that the answer to one question determines the next question that must be asked or answered. The field staff are carefully trained to follow the skip patterns but, under the current procedures, questions occasionally are missed and/or the wrong logical path is followed.

In the 1998 revision, the field staff will receive housing units to screnitiate, either from the listing instrument or from the Washington data base. The computer-assisted data collection instrument will receive the screnitiation schedules directly, bypassing the tedious and error-prone transcription process. It also will handle the skip patterns, so that only the appropriate questions are displayed during screening, and the housing units automatically receive the appropriate scope status. Because the computer is being used, the skip patterns can be much more efficient than before.

Because the computer has stored all of the previously collected data, automated logic checks can remove all redundant question patterns, thereby reducing the work of the field staff member and the respondent's burden. Assume, for example, that a field staff member asks about air conditioning equipment and the respondent says that there is a heat pump. From that one answer, the computer knows the air conditioning equipment, the heating equipment, and the heating fuel. The computer thus can "skip" all of the heating questions. Automated data checking will ensure that only correct data types are collected, other automated logic checks will ensure that collected data are consistent, and the instrument will inform the field staff member when all required data have not been collected. Because these data checks are being performed at the time of collection, errors and inconsistencies can be corrected while the respondent is present. The result is that the data that are sent to Washington will be as accurate as possible.

In previous revisions, if the field staff member determined that an address was in scope, he or she would once again transcribe housing unit identifiers to blank **initiation** forms.

The staff member would then collect the rent data and information on the specific housing services that were associated with the unit and the rent that was paid. This was done through another set of structured questions with the inherent skip patterns mentioned above. Once again, questions could be missed or the wrong path could be followed.

The computer-assisted data collection instrument will automatically flow into the initiation portion of the instrument when the instrument has determined that the screening is complete and the housing unit is in scope. As mentioned above, the collection instrument will handle the skip patterns, the automated data and consistency checks, the schedule completion checks, and the final initiation status. The screening and initiation data will then be electronically transmitted to the Housing data base in Washington.

In the current system, the screening and initiation data are shipped to Washington, where they are keyed by data capture staff and stored in the Housing data base. The analysts in Washington review and correct the data in preparation for the **pricing** of the housing unit. In the next on-panel month, a computer-generated pricing schedule for each in-scope housing unit is printed and shipped to the field. The field staff then, once again, collects data associated with the rent that is paid and the specific housing services that are associated with the unit. The pricing process is very similar to the initiation process, but some previous answers are provided on the pricing form. The collection of the housing data, and particularly the rent data, is independent. That is, the field staff member collects the data without giving the respondent the previous answer. Previous answers for some nonrent data are provided, so that the staff member can confirm certain changes with the respondent. These pricing data are shipped to Washington, where they are keyed, reviewed, and corrected as necessary. These data, along with the initiation or pricing data from 6 months earlier, are used in the Housing price relative calculation described below.

In the 1998 revision, the field staff will receive housing units to price from the Washington data base. The computer-assisted data collection instrument will automatically flow into the pricing portion of the process and, as mentioned above, will handle the skip patterns, the automated data and consistency checks, the schedule completion checks, and the final schedule status. The pricing data will then be electronically transmitted to the Housing data base in Washington.

On an infrequent basis, the scope status of the housing unit will be checked by repeating some of the questions asked during screening. This process may result in a change to the unit's scope status from in scope to permanently out of scope, temporarily out of scope, or incomplete. If this happens, the unit will be treated as described in the section above on screening.

In the 1998 revision, the Housing review and correction preprocessing system, which is described below, will handle the necessary preprocessing of the data obtained through computer-assisted data collection, and will be used in the review and correction of screening, initiation, and pricing data by CPI analysts.

Housing review and correction preprocessing system. The review and correction preprocessing system consists of various functions that must be performed before a unit can be used in index calculation. Upon receipt of the data, the system determines how to proceed (which functions are to be performed) based on selected variables, such as the scope status and the schedule status.

This system handles the micro data preprocessing required for the review and correction functions performed by the analysts. Some of these functions currently are performed as part of the existing price relative calculation (see below), so the analyst is not able to review all inputs to the calculation. As a result, the analysts occasionally have had to enter additional corrections, because the initial result of the calculation was not as expected. The new system also will complete all unit-level computations, so that the analyst will be able to review all micro data and price relative calculation inputs much earlier in the processing cycle. Once the calculation is run, the analyst will have to review only the results of aggregation and imputation.

The Housing review and correction preprocessing system will prepare all micro data necessary for the revised Housing price relative calculations for the 'rent' and 'rental equivalence' indexes. These computations include:

1. Derivation of normalized rents for units—These are the sums of the collected rents and the subsidy and work reductions, converted to monthly rents using the collected rent periods.

2. Determination and assignment of price relative calculation usability codes—These codes will be used by the Housing review and correction preprocessing system and in the price relative calculations for further computations for both the 'rent' and 'rental equivalence' indexes.

3. Calculation of utility adjustments required to adjust for changes in the utilities provided by the landlord.

4. Calculation of facility adjustments required to adjust for changes in other services provided by the landlord.

5. Calculation of aging bias adjustments used to adjust for the slight monthly depreciation of rental properties.

6. Calculation of structural change adjustments used to adjust for the addition to, or the removal from, the housing unit of central air conditioning, bedrooms, bathrooms, and other rooms.

7. Calculation of cost of utilities adjustments, which are needed to remove utility costs, if included, from the contract rents (used for the 'rental equivalence' index).

8. Calculation of economic rents and economic rent price relatives using the normalized rents (1 above) and the adjustments mentioned in 3, 4, 5, and 6 above. These economic rents will yield price changes of constant quality across time for the 'rent' index.

9. Calculation of pure rents and pure rent price relatives using the normalized rents (1 above) and the adjustments mentioned in 4, 5, 6, and 7 above. These pure rents will yield price changes of constant quality across time for the 'rental equivalence' index.

These data are then made available to the analysts through the review and correction instrument. This instrument is a subsystem of the review and correction preprocessing system and handles the interactive review and correction functions. The review and correction preprocessing system also permits the interactive derivation of dependently derived micro data, so that the analysts may immediately observe the results of their corrections.

Index estimation

The Housing and the Commodities and Services programs do not directly calculate indexes. Instead, they produce "price relatives," which are used in the index estimation system for basic index calculation. Price relatives are ratios of price change from the previous month ($T–1$) to the current month (T), and basic index calculation updates the last month's indexes ($T–1$) into the current month (T). (As explained above a decision was made for the Revision that the renter sample would be reweighted to represent owner units in the same segment.)

Weighting during the price relative calculation. The renter and owner costs of housing in the segment (see "Weighting during segment sample selection" above) become the basis of the renter and owner weights used by the price relative calculation for the segment.

To derived the total renter weight in the segment (RW_s), the segment weight (W_s) must be adjusted by the proportion of renter cost in the segment (RA_s) and the expectation of selecting a renter in the segment (RP_s).

The proportion of renter cost in the segment (RA_s) is the renter cost (RC_s) divided by the total cost (TC_s).

$$RA_s = RC_s / TC_s$$

The expectation of selecting a renter in the segment (RP_s) is the number of sampled housing units in the segment (SU_s) divided by the total number of housing units in the segment (HU_s).

$$RP_s = SU_s / HU_s$$

The total renter weight for the segment (RW_s), therefore, is the segment weight (W_s) times the reciprocal of the renter cost proportion (RA_s) adjusted by the proportion of sampled renters (RP_s).

$$RW_s = W_s * \frac{1}{RA_s * RP_s} = W_s * \frac{TC_s}{RC_s} * \frac{HU_s}{SU_s}$$

To derived the total owner weight in the segment (OW_s), the segment weight (W_s) must be adjusted by the proportion of owner cost in the segment (OA_s) and the expectation of selecting a renter in the segment (RP_s). Because owners are not being sampled and the renters are being reweighted to represent owners the RP_s is used in both derivations.

The proportion of owner cost (OA_s) is the owner cost (OC_s) divided by the total cost (TC_s).

$$OA_s = OC_s / TC_s$$

The total owner weight for the segment (OW_s), therefore, is the segment weight (W_s) times the reciprocal of the owner cost proportion (OA_s) adjusted by the proportion of sampled renters (RP_s).

$$OW_s = W_s * \frac{1}{OA_s * RP_s} = W_s * \frac{TC_s}{OC_s} * \frac{HU_s}{SU_s}$$

The renter and owner weights are ratios of expenditures, not expenditures themselves, so there is no need to convert them into quantities by dividing them by base rents or base implicit rents. In addition, the renter and owner weights are being derived from 1990 census data, while the first rent data will be collected no earlier than 1997, so there should be no autocorrelation effects. In short, there appears to be no fear of formula bias in the 'rent' and 'rental equivalence' estimators.

The 'rent' and 'rental equivalence' estimators. The 'rent' estimator is based on the change in the "economic rent," which is basically the "contract rent," adjusted for any changes in the quality of the housing unit. Because of the panel structure used in the Housing sample, the current economic rents for sampled, renter-occupied units within a segment, weighted by the renter weight, are divided by the previous ($T–6$) economic rents for sampled, renter-occupied units within a segment, weighted by the renter weight. The result represents the 6-month change in rent for all renter-occupied units in the segment.

In a parallel calculation, the current pure rents (which exclude the cost of any utilities included in the rent contract) for sampled, renter-occupied units within a segment,

weighted by the owner weight, are divided by the previous (*T*–6) pure rents for sampled, renter-occupied units within a segment, weighted by the owner weight. This is used as a proxy for the 6-month change in the equivalent rent for all owner-occupied units in the segment.

The functions of the price relative calculation have been designed to make use of the parallel 'rent' and 'rental equivalence' computations. In general, the calculation aggregates the weighted rents for the units (*i*) in the index area (*A*) for the current period (*T*) and for the period 6 months earlier (*T*-6), and then computes the price relatives:

$$REL_{T-6,T,A} = \frac{\sum_{i \in A} W_s * R_{i,T}}{\sum_{i \in A} W_s * R_{i,T-6}}$$

When the calculation is run for Rent, economic rents (ER_i) and renter weights (RW_s) are used. That is,

$$REL_{T-6,T,A}^{\text{rent}} = \frac{\sum_{i \in A} RW_s * ER_{i,T}}{\sum_{i \in A} RW_s * ER_{i,T-6}}$$

When the calculation is run for 'rental equivalence,'[4] pure rents (PR_i) and owner weights (OW_s) are used. That is,

$$REL_{T-6,T,A}^{\text{rental equivalance}} = \frac{\sum_{i \in A} OW_s * PR_{i,T}}{\sum_{i \in A} OW_s * PR_{i,T-6}}$$

Because the index estimation system needs a 1-month price relative, rather than a 6-month price relative, the 6th root of the $REL_{T-6,T,A}$ is derived:

$$REL_{T-1,T,A} = \sqrt[6]{REL_{T-6,T,A}}$$

and then passed to the estimation system for basic index computation for the 'rent' and 'rental equivalence' item strata.

The 'rent' and 'rental equivalence' item strata have not been changed in the 1998 revision item structure, but the revision area structure and the basic aggregation weights will be brought into the CPI in late 1997. The revision price relative calculation and the revision Housing sample will not be ready for use in the revised CPI until January 1999. During 1998, the current Housing sample will be priced, and the current price relative calculation will provide price relatives through December 1998. A Concordance program is being developed to map the price relatives for the current area structure to the revision area structure. □

Footnotes

[1] Primary sampling units are the metropolitan and nonmetropolitan areas defined as the CPI area sample. (See Janet L. Williams, "The redesign of the CPI geographic sample," pp. 10–17.)

[2] Eugene F. Brown and William H. Johnson, "Comparison of Stratification Designs for the Housing Sample of the Consumer Price Index," *1994 Proceedings of the Section on Survey Research Methods*, American Statistical Association.

[3] Two examples of out-of-scope addresses are units occupied by owners or relatives of the landlord. Because the owner sample is being dropped, owner-occupied units are out of scope. In the case of relatives, it has been determined that, if the tenant is a relative of the landlord, it is very difficult to collect data on market rents. Because the relative usually gets some type of rent reduction that cannot easily be determined, the units are considered out of scope.

[4] The price relative calculation also will handle the estimation of price relatives for the official Laspeyres index and the Geometric Means index. The same economic and pure rents, but different renter and owner weights, will be used for the Geometric Means index.