Application of Matrix Sampling Design on Inventory Estimation

Robin Lee, Susan Hinkins, and John Matson, Ernst & Young LLP
Robin Lee, Ernst & Young LLP, 1225 Connecticut Ave. NW, Washington DC. 20036

Key Words: Matrix Sampling Design, LIFO Inventory, Ratio Estimator, Replicates

Sampling can be used in many areas of business application. Whenever there are too numerous records or elements in the population to examine, statistical sampling can be used. Estimating the value or distribution of inventory in the retail industry is a good example. In this case, the population consists of retail stores and warehouses and the inventory items within these locations. For a retail chain with several thousand stores across the country, using the statistical sampling approach may be the only feasible solution as an alternative to a complete physical inventory of all stores.

The purpose of this paper is to introduce a particular sampling design that can be used to estimate the distribution of inventory across many categories under the sample size constraints. The category estimates are desired in proportion across the Bureau of Labor Statistics (BLS) Consumer Product Index (CPI) category. The CPI category proportion estimates are used to calculate the inflation or LIFO index for the tax return.

The sample is designed in December. The year end inventory data are collected in January, the end of the client's tax year. The Ernst & Young Quantitative Economics and Statistics (QUEST) group is responsible for designing the sample and estimating the CPI category proportions. The study has to be repeated annually.

Background

Until this sampling study was initiated in 1998, the client had been taking two samples – one for the CPI category proportion and the other for the cost to retail ratio (cost complement). To estimate the CPI category proportion, a random sample of stores was selected and all items carried in the stores were counted (wall-to-wall inventory). To estimate the cost complement, a two-stage sample was designed. The stores were sampled on the first stage and the items were subsampled within the sampled stores on the second stage.

In 1998, the client wanted to change the sample design so that both estimates can be calculated from one sample. In addition, the client wanted to do away with the wall-to-wall inventory because of the cost of hiring an outside inventory service provider to count all items in the sampled stores. The client wanted to redesign the two-stage sample so that it could be used for both sets of estimates. The client's preference was to sample many stores on the first stage and subsample only a few items within each sampled store on the second stage. The sample size of 30 was chosen as a reasonable number for the store staff members to count. The objective is to spread out the sample across more stores and limit each store sample to only a small number of items.

One design challenge is to provide reasonable estimates with a much reduced sample size. The number of items in the CPI categories far exceeds 30. If one considers the population as a matrix with stores as columns and CPI categories as rows, then the first stage sampling is a sample of the units of the matrix, hence the name of this sample design-- matrix sample design. The second stage sampling is to sample 30 items from the selected CPI category. It must be a fairly sparse design in that only a small proportion of the stores will be sampled -- 8%-9% of the stores. Assuming that similar items will be in the same part of the store, this design should have the advantage of further reducing the client's cost by assigning items from the same CPI category to a store for inventory

The point estimates, such as the total inventory value at retail in a particular CPI category, can be calculated by using standard probability weighting techniques. The challenge comes in estimating the variance of the point estimates, and more importantly, designing the sample so that the estimates of interest are reasonably precise. Although the estimates are desired at the CPI category level, the precision requirement is at the BLS Pool level. The pool is composed of many CPI categories. An example pool, Food and Beverages, is shown below to illustrate how the CPI categories in a pool get finer as the number of digits increases.

–01. Food and Beverages
–011. Food
–0111. Food at Home
–01111. Cereals and Bakery Products
–011111. Cereals and Cereal Products
–0111111. Flour and Prepared Flour Mixes
–0111112. Breakfast Cereal

–0111113. Rice, Pasta, Cornmeal
–01111131. Rice
–011112. Bakery Products
–0111121. Bread
–01111211. White Bread
–01111212. Bread Other Than White
–0111122. Fresh Biscuits, Rolls, Muffins
–0111123. Cakes, Cupcakes, Cookies

One other design challenge is that the sampling frame is not up to date. During the first year of sampling, many selected items were not found in the designated stores. In order to increase the probability of finding the sampled items, the item population was stratified into two group—core and non-core. The store population was also stratified into different size groups to create the replicates.

sample is also designed using the store replicates by stratifying the store population into three size groups—large, medium, and small—and randomly sampling one store (without replacement) from each stratum. Approximately 30 items are assigned to each store replicate to be inventoried; the same items were inventoried in each of the three stores. "Large" CPI categories are sampled more heavily, so that the CPI categories are sampled somewhat proportionate to size. The CPI category size is measured in terms of both the dollar amount and the number of items in each category.

Sample Design

Population

There are two population files. One is the list of all items stores carry, uniquely identified by Stock Keeping Unit (SKU) number. A drawback with this SKU frame is that it's not too current. The file contains too many items that are no longer carried in the stores. It is not easy to screen them out to drop them off from the frame before sampling. To help identify those items that are more likely to be found in the stores, the client puts two codes on the file. One denotes those core items that are carried by all stores and the other denotes the active items that have been ordered by the distribution center recently. The cross tabulation of the two variables showed that core items tended to be active and non-core items tended to be inactive. We label an item as 'primary' if it is either core or active. 'Secondary' status is assigned if the item is neither core nor active.

The other file is the list of all retail stores in business at the end of year. The store population provides, among other variables, year-to-date sales amount and the store size, which can be used to stratify the stores into three size groups.

Examples of important variables contained in the SKU population file are SKU number, item description, CPI code, unit cost, and unit retail price. The population parameters of interest from both store and SKU master files are summarized below.

- About 5,000 stores
- About 25,000 SKUs
- 130 plus CPI categories
- 6 Pools:
  - -Food & Beverages
  - -Household Furnishings
  - -Apparel
  - -Entertainment
  - -Personal Care
  - -Miscellaneous

2000 Sample Design

The stores are stratified into three strata based on the size of the store measured by sales area in square feet: small, medium, and large. The stratum boundaries are set to satisfy the Neyman allocation such that the product of the standard deviation (S_h) and the stratum count (N_h) is constant across all strata. The distribution of stores by strata is:

Large:	1,836
Med:	2,172
Small:	944

A store replicate consists of three stores. Each replicate is formed by randomly choosing one store (without replacement) from each size stratum. From the population of 4,955 stores, 432 stores were sampled to create 144 store replicates.

A SKU replicate consists of 30 items. Using the item label, 29 primary SKU's and 1 secondary SKU are assigned to each SKU replicate.[1] Ideally, one would like to confine attention to the primary items; this would greatly simplify the sampling and the estimation process. However, one cannot assume that the secondary items are no longer in the stores. We assume, however, that they are less likely to be

[1] The lopsided sample allocation was one of major shifts in the population in year 2000. Using the Core and Active item indicator from the client, 97% of the SKU population was determined to be primary. The 1999 sample had 20 primary and 10 secondary in each SKU replicate.

in the stores and therefore sample the secondary items at a much smaller sampling rate than the primary items. The information returned is the number of sampled items found in the inventory for each SKU on the list.

The 2000 sample had to be designed to estimate 130 CPI categories. The sample was designed so that estimates of smaller subdivisions of CPI categories could be made while keeping the sample size small. Some CPI categories are very small - containing very few prima ry SKU's. For sampling purposes, the very small CPI categories are grouped together to form CPI groups that contain at least 0.3% of the inventory. This reduces the 130 CPI categories to 30 CPI groups. The CPI categories are grouped to combine similar types of items into the same group, to facilitate the inventory process. Categories are never combined across "Pools." Replicates are assigned to CPI groups.

In order to use Generalized Variance estimates, super-strata are defined for allocation of the replicates to the CPI groups. Sufficient replicates should be assigned to each super-stratum so that stratum variances can be reasonably estimated. Ideally, 8-10 super-strata of approximately the same size would be defined, where "size" is defined below. However, this is not possible because each super-stratum must be contained within a pool. Some Pools are very small, but they must each be kept as separate strata. Therefore, the small pools are kept as a separate super-strata. The number of replicates assigned for each pool is shown in the table below.

Table 1. Definition of 11 Super-Strata/Pool

Pool	CPI groups	Size	Additional Replicates	Total # of Replicates
1	1-6	9.9	11	17
2	7	8.2	9	10
2	8	9.8	11	12
2	9, 10	9.9	12	14
2	11 - 15	9.8	11	16
3	16, 21 – 23	11.	13	17
3	17	10	12	13
3	18-20	11.3	13	16
4	24 – 26	7	8	11
5	27 – 28	9.5	11	13
6	29 - 30	3	3	5
Total				144

Estimation

The data returned from the inventory consist of the number of items found in each sample store, for each designated SKU. A retail cost or value is associated with each SKU so that the information can be converted to a dollar value for each sampled SKU. Each replicate corresponds to a CPI category or a CPI group. That is, the SKUs designated for a replicate are a sample of the SKUs in a particular CPI category or group.

For sample selection, the SKUs were categorized using two indicators. The indicator Core_Flag has three possible values, and the value of interest in this case was 'C' indicating a "core" SKU. There was also an indicator of "active" vs "inactive." A SKU is classified as "active" if the Last_date is July 1, 2000 or more recent than July 1, 2000. It was hypothesized that SKUs that were neither "active" nor "core" were items that had been in the inventory at some time but were either no longer in the inventory or were only residually in stores. But because this is not known, these SKUs could not be eliminated from the sampling population. Each replicate was defined as containing 29 SKUs that were either core and/or active and 1 SKU from the population of SKUs that were neither active nor core.

The sample results indicate that this classification of SKUs was not effective. Over all CPI codes, 31% of the sampled active/core SKUs were found in the stores, and 29% of the inactive/noncore SKUs were found. Therefore, overall, the classification used to try to predict which SKUs were not likely to be in the stores, failed. When the rates are compared by CPI group, one finds that for most groups, there is little difference between the two types of SKUs in terms of the relative frequency of finding the SKUs in the stores. For some SKUs the "response" rate for the core/active SKUs is higher, but for a few CPI groups the "response" rate is lower for the core/active SKUs. The following table shows the relative frequency with which sampled SKUs were found in the stores, by their classification as core and active. From this sample of SKUs it would appear that the best indicator would be core vs non-core where a core SKU is defined as having the Core_Flag equal to 'C'.

Table 2. Percentage of Selected SKUs Found in the Sampled Stores

	Over All	Active	Not "Active"
Core	64.8%	66.7%	53.3%
Non-Core	24.1%	23.2%	25.0%

But because the non-active, non-core SKUs are found in the stores (25% were found), this also indicates that one cannot exclude SKUs based on these indicators. Stratifying directly on the indicators can result in extreme values for weights. In the future it might be advisable to allocate the SKUs proportionally by the "core" vs "non-core" status to ensure that "core" SKUs are included in the sample in each replicate. (In one CPI category, the 2000 sample was unlucky in not selecting any SKUs in the "core" category.) But the estimation could still be done as a post-stratified estimate using the CPI group only.

The estimation for the year 2000 was done by post-stratifying to the CPI group total, rather than using the core/active designation. This is described below.

<u>Notation</u>. For a particular CPI category or group, say g, there are M_g SKUs in the population. We treat the sample as if in each replicate, m_g =30 SKUs were randomly selected from the M_g SKUs. Even though we sampled by selecting from core/active and non-core/inactive categories of SKUs within CPI groups, we can calculate the estimates by combining results to the replicate/group level. We may sacrifice some precision in the estimates this way, but by combining the data, reasonably stable standard errors can be estimated. As a check, point estimates were calculated using the weights consistent with the design, and the estimated proportions essentially match the estimates shown in Tables 1 and 2.

Let x_{ghri} denote the total dollar value for the ith SKU returned for the store sampled in stratum h, in replicate r for group g, and let $x_{ghr.}$ denote the total dollar value of the m_g SKUs returned for the sampled store in stratum h, in replicate r, which corresponds to group g. For example, five replicates were assigned to CPI category 111512; therefore for this CPI group g, r =1,...5. In each replicate there will be data from 3 stores (h=1,2,3).

Estimation. The quantity to be estimated is the proportion of the total dollar value of the inventory that is contained in each CPI category. Although the assignment of the number of replicates to CPI categories sometimes used a random assignment with pps, the estimation procedure is conditional on the number of replicates assigned and does not incorporate the random aspect in allocation of replicates.

The estimate of the total inventory in group g is calculated as:

$$\hat{X}_g = \frac{1}{R_g}\sum_{r=1}^{R_g}\frac{M_g}{m_g}\sum_{h=1}^{3}N_h x_{ghr.}$$, where R_g denotes the total number of replicates assigned to group g.

In this design, m_g=30 for all groups, all replicates, which simplifies the calculation somewhat.

The estimated percentage of the inventory in group g is then calculated as

$$\hat{P}_g = \frac{\hat{X}_g}{\sum \hat{X}_j}$$ where the denominator is the sum over all groups.

Estimates for a CPI code, C, within a CPI group, are calculated similarly, where the summation of SKUs within the replicate is over only those SKUs that are in the CPI category of interest:

$$\hat{X}_C = \frac{1}{R_g}\sum_{r=1}^{R_g}\frac{M_g}{m_g}\sum_{h=1}^{3}N_h \sum_{j\in C} x_{ghrj}\ .$$

The estimated percentage of the inventory in group g is $\hat{P}_g = \frac{\hat{X}_g}{\sum \hat{X}_j}$ where the denominator is the sum over all groups.

<u>Variance estimation</u>. A random groups estimate is used for variance estimation (Wolter, 1985). The sample is divided randomly into 10 smaller samples, within the original design. In each replicate, two Primary SKUs and one Secondary SKU are randomly selected (without replacement) and assigned to each random group. In this way, we have 10 samples (each sample is approximately 1/10 the size of the original sample) and 10 sample estimates, where the estimate is made as described above. Let $\hat{P}_{gk}$ indicate the estimate for CPI Group g, calculated from the data in random group k. The random group variance estimate for the original estimate $\hat{P}_g$ is constructed as

$$v_2 = \frac{1}{10(10-1)} \sum_{k=1}^{10} (\hat{P}_{gk} - \hat{P}_g)^2$$

where there are 10 random groups.

This estimator is not linear; the mean of the 10 random group estimates is not the same as the original estimate described above: $\hat{P}_g \neq \bar{\hat{P}}_g = \frac{1}{10} \sum_k \hat{P}_{gk}$. However, one expects that $\hat{P}_g \doteq \bar{\hat{P}}_g$, and, if not, then one may not feel comfortable using this variance estimator. Using the original data, the two estimators were not always "close enough" to feel comfortable with the estimates. Therefore statistical outliers were identified and the weights were reduced, as described below.

Smoothing Outliers. Outliers may indicate data errors (2 were verified in the sample data) or they may indicate statistical outliers. Each random groups estimate is based on a relatively large sample size, in terms of the number of stores. Therefore, one would expect that the random groups estimators should be approximately normal. As a check on the data quality, we compared the random groups estimates to the normal distribution using qq-plots. This identified certain outliers. In these cases, the sample weights for the secondary SKUs were adjusted so that the estimators more closely followed the normal distribution.

Estimates. The resulting estimates and the standard errors are shown in the following two tables. The next table gives the estimates by CPI "pool" where the CPI categories were pulled together based on the first digit of the CPI codes. The point estimates for the CPI codes are not shown here.

Table 3. 2000 Year End Inventory Estimates by pool

Pool	$\hat{P}$	Standard Error for $\hat{P}$
1. Food and Beverage	.0481	.0041
2. Household Furnishings	.3406	.0120
3. Apparel	.4378	.0164
4. Entertainment	.0743	.0074
5. Personal Care	.0616	.0061
6. Miscellaneous	.0377	.0089

REFERENCES

Cochran, William (1977), *Sampling Techniques*, John Wiley & Sons

Wolter, Kirk (1985), *Introduction to Variance Estimation*, Springer-Verlag