Evaluating Item Rotation for the New Goods Initiative

Economists in research agencies and other think-tanks have long argued that the Consumer Price Index (CPI) has been slow to bring newer items into the sample. Another criticism is that the distribution of goods and services in our sample doesn't reflect, well-enough, the distribution of goods and services in the U.S. The New Goods Initiative was developed to address these kinds of concerns.

In March of 2004, the Item Rotation Evaluation Team (IREVT) delivered the Final Report on evaluating Within Outlet Item Rotation (IR) to the Commodities and Services Steering Committee (CSSC). The objectives for IREVT were to determine if IR was a success, supply the CPI program with numerous characteristics about the IR sample, and deliver tools and methods for continual evaluation. Future evaluations will be performed by the C&S Data Collection Review and Improvement Team (CRIT).

This article will present some of the content from that report. It will provide background to item rotation; explain our basic discoveries about the sample; and present our methodology and ideas for future evaluation.

Background

Item rotation was pioneered in ELIs that used Directed Substitution, and was also tested using prescription drugs. It is a way to include in the CPI a new set of unique items selected through disaggregation that reflect more up-to-date sales volume. The new items include new goods or services that potentially did not even exist when the original sample was first initiated, including new goods captured under a newly revised Entry Level Item (ELI) definition.

Stated more simply, items will be replaced in existing outlets *halfway* through their original initiation cycle. This effort will not add more quotes to the Index, because all of the new quotes are *replacements*. Because outlets are rotated every *four* years, the goal is to rotate the quotes within those outlets every *two* years (hence the term "within outlet").

The first kind of item rotation uses the *Initiation in Pricing* method: new quotes from the same ELI are added to existing pricing outlets, by replacing some or all of the current quotes. Then, the *replaced* quote drops out of the outlet a month after the end of the IR initiation window.

The second kind of item rotation uses the *Reinitiation with Comparison* method: existing quotes are simply reinitiated in the outlet. System generated messages for the selected quotes are sent to data collectors in the field, letting them know it is time to reinitiate.

Evaluation

Our team had a chance to evaluate 22 ELIs. The IR sample was fielded in 2003 using the initiation sample fielded in February of 2001. This was the first round of item rotation. The Initiation in Pricing method rotated 643 quotes in 18 ELIs, across 38 PSUs. The Reinitiation with Comparison method rotated 630 quotes in 8 ELIs, across 18 PSUs (we were not able to evaluate some ELIs due to seasonality).

The results of this evaluation for the first round of item rotation, based on the *initiation* sample sent to the field in February of 2001 can be found in Table 1.

Table 1: Generalized Evaluation

ELI Evaluation:

- Number of ELIs: 22
- ELIs receiving positive ratings: 5 (22.7%)
- ELIs receiving negative ratings: 3 (13.6%)
- ELIs receiving undetermined/no-change ratings: 14 (63.6%)

Cluster Evaluation:

- Number of clusters which had specific expectations: 45
- Clusters receiving positive ratings: 12 (26.7%)
- Clusters receiving negative ratings: 9 (20%)
- Clusters receiving undetermined/no-change ratings: 24 (53.3%)

The team made some interesting discoveries after examining this item rotation sample. For personal computers, we found that Directed Substitution was serving its designated purpose and should be continued, but results from Reinitiation with Comparison for personal computers were not as good as we hoped. For prescription drugs, the number of uninitiated quotes increased 200%, indicating a potential concern with respondent burden. However, item rotating prescription drugs does successfully bring newer drugs into the sample (i.e., prescription drugs that are newer than they would be under outlet-only rotation).

Excluding personal computers and prescription drugs, we discovered that item rotation has no impact on outlet refusals and negligible impact on canceling quotes. For the overall sample, there is plenty of time to do both initiations and reinitiations. Also, the majority of the quotes analyzed were disaggregated using Percent of Sales or Ranking, and at least 3 steps in disaggregation. Finally, the team discovered that IR had relatively little bearing on the number of visits required to collect the quotes.

The team also answered some interesting questions. One question asked was "Is the disaggregation process successful in obtaining a distribution of goods that better reflects the marketplace?" While the CPI database serves as a useful source of information about the goods and services in our samples, we aren't aware of any central marketplace data about goods and services in the U.S. as a whole. This makes it difficult to provide an answer to this question. Instead, the team compared & contrasted specification elements deemed important quality factors by commodity analysts (there's too much detail about this to discuss here; the detail was included in the appendices of the Final Report).

Another interesting question asked of the team was "Is the disaggregation process successful in bringing new goods into the Index?" The team realized we did not have enough information to answer this question for any ELI except prescription drugs. For prescription drugs, we used the age of the drug to indicate whether it was a new item. For the other ELIs, we would have to physically look at each quote's checklist specifications. Therefore, in an effort to get the answer to this question without having to do this, we started to develop a detailed output tool (described later in this article); it is not currently part of our methodology.

Methodology

Besides the discoveries we made about item rotation, the overall methodology involved in evaluation is also worth mentioning. The methodology is a five step process: determine the characteristics of the IR sample, gather and record commodity analysts' expectations, create and distribute computer program diagnostic output to commodity analysts, record commodity analysts' evaluations, and report the results. Most of the process centers around creating the computer program diagnostic output, and recording the commodity analysts' evaluations.

The methodology for evaluating item rotation samples starts with commodity analysts in OPLC because they have the necessary skill, education, and experience. Therefore, the initial and final steps in the methodology generally involve human intelligence, while the steps in between involve computer intelligence. The steps in between use one computer program for extracting the data from the mainframe database, and another one for creating the reports.

Then, the computer program diagnostic output is presented to commodity analysts, allowing them to use it to make a generalized evaluation of item rotation. Success was first evaluated based on whether item rotation met expectations for the ELI as a whole. Then, success was evaluated based on whether expectations for each individual cluster were met (i.e., instances where the quotes were expected to move into or out of the cluster, or instances where changes in specification distributions were expected).

Commodity analysts gave each ELI or cluster a positive, negative, or undetermined/no-change rating. A positive rating meant that the expectations were met. A negative rating meant that the expectations were not met. A rating of undetermined/no-change meant that no discernable differences could be observed in the sample after item rotation.

Ideas for the Future

Forming generalized evaluations in the manner described above is a highly subjective process; sometimes we may want to be more specific. For example, as mentioned earlier in this article, the team did not have enough information to ascertain whether the disaggregation process placed newer goods into the sample. However, using a detailed output tool in our methodology could help commodity analysts answer this question quickly and automatically; without the time-consuming manual work usually needed for such a task.

A detailed output tool, like the one the team worked on, can provide more specific information. The tool does this by storing information obtained from commodity analysts into its memory; definitions of new items for each cluster, and expectations about those items. Then, the program uses this information at execution to decide whether or not expectations were met for each quote it evaluates. Each quote is displayed with a complete set of specifications. This is illustrated in condensed format in Table 2.

Table 2: Example of Detailed Output Tool Results

Quote Code	ELI	Cluster	1st Specification Expected	2nd Specification Expected	1st Found	2nd Found	Reject/Accept
003	EE021	01A	A2 - SOFTWARE	B1 - FULL VERSION	A2	B1	ACCEPT
001	EE041	01B	A1 - HOME TEL	B6 - CORDLESS	A1	B7	REJECT
110	FJ041	01A	Unable to elaborate		A1		Inconclusive

Evaluation of the hypothesis for each quote is either: Accept, Reject, Inconclusive, or some other set of values if desired. For example, information about each quote's relative importance or ELI/cluster distribution could be plugged into the program. Then, the decision rendered by the tool is more meaningful: a quote with a high relative importance could have a weight of 5.00 while a quote with a low relative importance could have a weight of 0.50. Then the 1st quote = ACCEPT x 5.00, while the 2nd quote = ACCEPT x 0.50, giving each quote its own weighted decision. Ultimately, some rejected null hypotheses will become more important than others.

Information produced by the detailed output tool is furnished by applying it to both the old and the new samples. The tool gives a percentage of new items for any sample it is run against; therefore, running it against both samples produces two percentages.

The results are placed into an Excel spreadsheet. Then, simple Excel formulas turn those two percentages into one single percent (positive or negative) to measure the new item success for that round of item rotation. Table 3 illustrates the results for the first round of item rotation.

Table 3: Detailed Output Tool - New Items Found

2003 IR New Sample		**2001 Initiation Old Sample**	
Total quotes in the sample	643	Total quotes in the sample	643
(Less) deleted quotes	(2)	(Less) deleted quotes	(2)
(Less) medical quotes	(454)	(Less) medical quotes	(454)
(Less) uninitiated quotes	(16)	(Less) uninitiated quotes	(0)
(Less) quotes w/o expectations	(27)	(Less) quotes w/o expectations	(28)
Subtotal	*144*	*Subtotal*	*159*
New Items Found	34	New Items Found	28
(Null Hypothesis was REJECTED)		(Null Hypothesis was REJECTED)	
Total Eligible Quotes	÷ 144	Total Eligible Quotes	÷ 159
Percentage of New Items Found	0.236	Percentage of New Items Found	0.176

Formula

$$C = \frac{x_2 - x_1}{x_1}$$

C = relative change

x_1 = initial value

x_2 = final value

$$C = \frac{34 - 28}{28} = 21.43$$

The output illustrated in Tables 2 and 3 is then delivered in a spreadsheet to commodity analysts, to supplement their evaluation.

A tool like this one could be an improvement to our methodology for evaluating item rotation. Along with input from commodity analysts, it could potentially be the beginning of an evaluation process for new items that is feasible, not time-consuming, quantifiable, and well documented.

Besides being able to answer more specific questions about new items, there is another improvement that would make evaluating item rotation easier: access to more information. For example, it would be useful to have some measure of the *distribution* of unique items in the old sample, to have at least some *a priori* expectations about the condition of the old sample's distribution of unique items vs. that of the new sample. It would also be useful to have some measure of the distribution of goods in the marketplace as a whole. By this, we mean the distribution of expenditures on goods and services in the U.S. economy, and how they relate to what we have in our price index samples.

Conclusion

In the end, the entire evaluation process was quite fascinating and very informative. Ultimately, we hope the information, strategies, and methods presented in our report can help us better understand the CPI items we price each and every day.