COMPOSITE ESTIMATION IN THE INTERNATIONAL PRICE PROGRAM

Louis Rizzo, Bureau of Labor Statistics

The International Price Program (IPP) publishes quarterly indexes of import and export price levels in Standard International Trade Classification (SITC) product categories. These indexes are fixed-weight Laspeyres indexes, with the fixed weights based on quantity traded in the designated base quarter. To estimate these indexes samples are taken periodically of importers and exporters in each product category. To estimate the index, the sample needs to be longitudinal: product areas within the company are followed quarterly over a long period. Since international trade tends to be volatile, a given sample can deteriorate markedly as sampled product lines drop out of trade. Therefore a new sample is needed at least every two years to support the index over time. At any given time, however, there will be multiple samples generating data for the index. Each sample generates an estimate of the index: it is necessary to find a way to put together the sample estimates into a composite estimator of the index. This paper describes one way of doing this. The first section describes the estimation problem. The second section then describes the estimation methodology and gives an example of its use. The third section describes the statistical qualities of the estimator.

I. BACKGROUND

The formal fixed-weight Laspeyres index for a given SITC category in quarter T will be

$$I_T = \frac{\sum_{i=1}^{N} P_{i,T}\, q_{i,o}}{\sum_{i=1}^{N} P_{i,o}\, q_{i,o}} \qquad (1)$$

where $q_{i,o}$ is the quantity traded of a given item i of the market basket in the base quarter, $P_{i,o}$ is the price level of that item in the base quarter, $P_{i,T}$ is the price level of the item in quarter T and N is the number of items in the market basket in the population.

There are relatively exact measures of $P_{i,o}\, q_{i,o}$ in the base period from U.S. Bureau of the Census and U.S. Customs data on trade revenues.

Let

$$W_{i,o} = P_{i,o}\, q_{i,o} \quad \text{and} \quad R_{i,T,o} = \frac{P_{i,T}}{P_{i,o}}.$$

$W_{i,o}$ is the total dollar value of trade of item i in the base quarter and $R_{i,T,o}$ is the long term relative of item i between time 0 and time T. Then I_T can be rewritten:

$$I_T = \frac{\sum_{i=1}^{N} R_{i,T,o}\, W_{i,o}}{\sum_{i=1}^{N} W_{i,o}} \qquad (2)$$

I_T can be rewritten as a product of short-term price level changes,

$$\text{Let } F_t = \frac{\sum_{i=1}^{N} R_{i,t,o}\, W_{i,o}}{\sum_{i=1}^{N} R_{i,t-1,o}\, W_{i,o}} \qquad (3)$$

where t=1, 2,....,T.

Then from algebra

$$I_T = \prod_{t=1}^{T} F_t \qquad (4)$$

(see [1])

Each F_t is the measure of the price level change from quarter t-1 to quarter t based on the fixed market basket of quarter 0 (technically called the short-term-relative). These are the short-term price level changes, and the overall index I_T can be written as a product of these short-term changes.

To estimate I_T, a sample is taken in the base quarter of all items: each item i receives a probability of selection $\pi_{i,o}$ based on its weight $W_{i,o}$. Let n_1 be the sample size. Then the Horvitz-Thompson estimator of I_T from sample 1 is

$$\hat{I}_{T,1} = \frac{\sum_{i=1}^{n_1} \frac{W_{i,o}}{\pi_{i,o}} R_{i,T,o}}{\sum_{i=1}^{n_1} \frac{W_{i,o}}{\pi_{i,o}}} \qquad (5)$$

(For discussion of this estimator, see [5]).

$\hat{I}_{T,1}$ is slightly biased, having the well-known bias of a ratio estimator. Ideally each sample unit i will send in data in every quarter t for a six year period. Then estimates $\hat{I}_{T,1}$ can be calculated for T=1,....24. In practice, however, many sample units do not or cannot respond in a given quarter, or drop out all together because of changing trade patterns.

Therefore every product category is resampled every other year (i.e. in quarter 8, 16, etc.). A sample (designated here as sample 2) is taken again over all items: each item i receives a probability of selection $\pi_{i,8}$ based on its 8th quarter revenue $W_{i,8}$. Let n_2 be the sample size in the product area for sample 2. Then there is a second estimator of I_T for T > 8:

$$\hat{I}_{T,2} = \frac{\sum_{i=1}^{n_2} \frac{W_{i,8}}{\pi_{i,8}} R_{i,T,8}}{\sum_{i=1}^{n_2} \frac{W_{i,8}}{\pi_{i,8}}} \qquad (6)$$

Another full sample (sample 3) is taken two years later in quarter 16. This sample generates an estimate of I_T for T > 16:

$$\hat{I}_{T,3} = \frac{\sum_{i=1}^{n_3} \frac{W_{i,16}}{\pi_{i,16}} R_{i,T,16}}{\sum_{i=1}^{n_3} \frac{W_{i,16}}{\pi_{i,16}}} \qquad (7)$$

By the next sampling period (24), the first sample begun in period 0 is discontinued, so in any given quarter t after the 16th quarter, there are three estimators of I_T. Each estimator is of the full index, based on a different sample of a slightly different market basket. A methodology was needed to bring together the various sample estimates into a composite estimator of I_T for all T. This problem is very similar to the problem of combining estimates from multiple frames discussed in (for example) the papers of Cochran, Hartley, and Lund.([2], [4], and [6])

II. THE ESTIMATION METHODOLOGY

There is a simple theory for bringing together three estimates $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\mu}_3$ of a given parameter μ. If in fact $E(\hat{\mu}_1) = E(\hat{\mu}_2) = E(\hat{\mu}_3) = \mu$, then we can choose a 'best' estimate among among the set of all linear combinations of $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\mu}_3$. Let $\underline{\hat{\mu}} = [\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3]'$. Then the minimum variance unbiased estimator (BLUE) of μ with the form $\underline{W}'\,\underline{\hat{\mu}}$ is found by setting $\underline{W}'$ equal to

$$\underline{W}'^{*} = (\underline{1}' \textstyle\sum_{\mu}^{-1} \underline{1})^{-1}\, \underline{1}' \textstyle\sum_{\mu}^{-1} \qquad (8)$$

where $\sum_{\mu}$ is the variance matrix of $\underline{\hat{\mu}}$. A derivation of this can be found in [8], p. 31.

The methodology chosen was to focus on finding an efficient estimator for each short-term relative F_t. Once an $\hat{F}_t$ was found for each quarter t, then

$$\hat{I}_T = \prod_{t=1}^{T} \hat{F}_t \qquad (9)$$

Therefore F_t is the parameter that is being estimated by the three samples active in quarter t. Let these three sample estimates be

$\hat{F}_{t,1}$, $\hat{F}_{t,2}$, $\hat{F}_{t,3}$. Then suppose

$\underline{\hat{F}}_t = [\hat{F}_{t,1}, \hat{F}_{t,2}, \hat{F}_{t,3}]'$. If $\sum_{F,t}$ is the variance matrix of $\underline{\hat{F}}_t$ then the BLUE of F_t is

$$\hat{F}_t = (\underline{1}' \textstyle\sum_{F,t}^{-1} \underline{1})^{-1}\, \underline{1}' \textstyle\sum_{F,t}^{-1} \underline{\hat{F}}_t \qquad (10)$$

Of course, $\sum_{F,t}$ will never be known in practice. At best, we will be able to assume the covariance terms are zero under certain sample designs (i.e. those sample designs where the samples are independent). The variances will never be known, but need to be estimated. Any of a number of methodologies can be used such as random groups, balanced half sampling, model-based techniques, Taylor series, etc.

When we cannot assume the covariances are zero, then we also need estimates of $\mathrm{Cov}(\hat{F}_{t,i}, \hat{F}_{t,j})$, the non-diagonal elements of $\sum_{F,t}$. In both cases, the estimator of F_t will be

$$\hat{F}_t = (\underline{1}' \hat{\textstyle\sum}_{F,t}^{-1} \underline{1})^{-1}\, \underline{1}' \hat{\textstyle\sum}_{F,t}^{-1} \underline{\hat{F}}_t = \underline{\hat{W}}_t'\, \underline{\hat{F}}_t \qquad (11)$$

An Example

An example of an actual index calculated by these methods is given below. The index measured is the price level of imported alcoholic beverages, not including beer and wine. There is an ongoing sample, sample 1, and a new sample that is fielded in February, 1983. It takes a full year to initiate every sample unit, so sample 2's initiation is completed by February, 1984. The quarterly short-term-relatives estimated by the two samples is given below:

Table I

Quarter	$\hat{F}_{t,1}$ Sample 1 STR	Sample 1 % Change	$\hat{F}_{t,2}$ Sample 2 STR	Sample 2 % Change
1 1983	1.08071	+ 8.07%	-	-
2 1983	.994451	- .55%	-	-
3 1983	1.01809	+ 1.81%	.99181	- .82%
4 1983	.991241	- .88%	.99950	- .05%
1 1984	1.00924	+ .92%	1.00747	+ .75%
2 1984	1.00375	+ .38%	1.00242	+ .24%
3 1984	1.00147	+ .15%	1.00462	+ .46%
4 1984	.998721	- .13%	.98456	- 1.55%
1 1985	1.00396	+ .40%	.99819	- .18%
2 1985	1.00247	+ .25%	1.01355	+ 1.35%
3 1985	1.00844	+ .84%	1.00954	+ .95%
4 1985	1.01728	+ 1.73%	1.01274	+ 1.27%

Variance estimates were calculated for each sample STR (a model-based technique was chosen in this case, but any technique can be used).

The first column in Table II is the estimated variance of the short term relative estimated by sample 1 in the quarter. This column is followed by the estimated panel weight for sample 1's estimate based on the variance and covariance estimates and equation 10. The third column is the sample size: the number of responding sample units for sample 1.

The variance estimate, panel weight, and sample size for sample 2 are given in columns 4 through 6 of Table II. Column 7 is the estimated covariance (based on a model-based estimator also) between the panels, and the last column is the number of overlapping sample units.

Table II

Quarter	$v(\hat{F}_{t,1})$	$w_{t,1}$	n_1	$v(\hat{F}_{t,2})$	$w_{t,2}$	n_2
1 1983	8.6×10^{-4}	1.000	53	-	-	0
2 1983	1.4×10^{-4}	1.000	48	-	-	0
3 1983	5.2×10^{-4}	.395	47	3.4×10^{-4}	.605	61
4 1983	2.5×10^{-4}	.344	49	1.7×10^{-4}	.656	85
1 1984	1.8×10^{-4}	.267	49	$.9 \times 10^{-4}$	.733	105
2 1984	2.0×10^{-4}	.247	48	1.0×10^{-4}	.752	112
3 1984	0.8×10^{-4}	.231	49	$.4 \times 10^{-4}$	.769	113
4 1984	7.8×10^{-4}	.254	51	3.9×10^{-4}	.746	116
1 1985	3.4×10^{-4}	.262	49	1.7×10^{-4}	.738	108
2 1985	2.5×10^{-4}	.265	46	1.3×10^{-4}	.735	104
3 1985	3.8×10^{-4}	.281	51	2.0×10^{-4}	.719	108
4 1985	3.7×10^{-4}	.274	49	1.9×10^{-4}	.726	108

Quarter	$c(\hat{F}_{t,1}, \hat{F}_{t,2})$	n_{12}
1 1983	-	0
2 1983	-	0
3 1983	1.5×10^{-4}	17
4 1983	$.7 \times 10^{-4}$	17
1 1984	$.4 \times 10^{-4}$	18
2 1984	$.5 \times 10^{-4}$	18
3 1984	$.2 \times 10^{-4}$	19
4 1984	1.4×10^{-4}	19
1 1985	$.8 \times 10^{-4}$	19
2 1985	$.6 \times 10^{-4}$	19
3 1985	$.9 \times 10^{-4}$	19
4 1985	$.9 \times 10^{-4}$	19

If there was no covariance, then the panel weights would be inversely proportional to the estimated variance. With positive covariance, the weaker panel is given an even lower panel weight than its variance would indicate alone.

Table III gives the the calculated index estimated by each sample alone, the composite short term relatives, and the composite index.

Table III

Quarter	Composite STR Estimate and Index Estimate $\hat{F}_t$	$\hat{I}_t$	Single Sample Index Estimates $\hat{I}_{t,1}$	$\hat{I}_{t,2}$
1 1983	1.081	1.0867	1.0867	-
2 1983	.994	1.0747	1.0747	1.0747(I)*
3 1983	1.002	1.0771	1.0941	1.0659
4 1983	.998	1.0746	1.0878	1.0654
1 1984	1.008	1.0831	1.0979	1.0733
2 1984	1.003	1.0861	1.1020	1.0759
3 1984	1.004	1.0903	1.1036	1.0809
4 1984	.988	1.0774	1.1022	1.0642
1 1985	1.000	1.0771	1.1066	1.0623
2 1985	1.011	1.0885	1.1093	1.0767
3 1985	1.009	1.0986	1.1187	1.0869
4 1985	1.014	1.1139	1.1380	1.1008

The second quarter, 1983 value of $\hat{I}_{t,2}$ is imputed based on the second quarter index estimate from the first sample. $\hat{I}_t$ ends up between the two separate sample index estimates $\hat{I}_{t,1}$ and $\hat{I}_{t,2}$. This would be expected most of the time.

III. STATISTICAL CHARACTERISTICS OF THE ESTIMATOR

If each sample short term relative estimate $\hat{F}_{t,i}$ is biased, then the overall index estimate $\hat{I}_T$ may have a significant bias. We can write $\hat{F}_t$ as:

$$\hat{F}_t = F_t + \gamma_t$$

where F_t is the true population short term relative and γ_t is the residual error term of the composite estimator. Then

$$\hat{I}_T = \prod_{t=1}^{T} (F_t + \gamma_t)$$

$$= (F_1 + \gamma_1)(F_2 + \gamma_2) \cdots (F_T + \gamma_T)$$

$$= \prod_{t=1}^{T} F_t + \sum_{t=1}^{T} \gamma_t \prod_{k \neq t}^{T} F_k$$

$$+ \sum_{t=1}^{T} \sum_{\substack{s=1 \\ s \neq t}}^{T} \gamma_t \gamma_s \prod_{\substack{k \neq t \\ k \neq s}}^{T} F_k + \cdots + \gamma_1 \gamma_2 \cdots \gamma_T \tag{13}$$

Since $I_T = \prod_{t=1}^{T} F_t$, by dividing both sides by $I_T = \prod_{t=1}^{T} F_t$ we obtain:

$$\frac{\hat{I}_T - I_T}{I_T} = \sum_{t=1}^{T} \frac{\gamma_t}{F_t} + \sum_{t=1}^{T} \sum_{\substack{s=1 \\ s \neq t}}^{T} \frac{\gamma_t \gamma_s}{F_t F_s}$$

$$+ \cdots + \frac{\gamma_1 \gamma_2 \cdots \gamma_T}{F_1 F_2 \cdots F_T} \tag{14}$$

The expectation of the left side of the above equation is the rel-bias of $\hat{I}_T$.

$$\text{rel-bias }(\hat{I}_T) = E\,\frac{\hat{I}_T - I_T}{I_T} = \sum_{t=1}^{T} \frac{E(\gamma_t)}{F_t} + \sum_{t=1}^{T} \sum_{\substack{s=1 \\ s \neq t}}^{T} \frac{E(\gamma_t\, \gamma_s)}{F_t\, F_s} + \cdots + \frac{E(\gamma_1\, \gamma_2 \cdots \gamma_T)}{F_1\, F_2 \cdots F_T} \qquad \text{(eq 16)}$$

If each sample short term relative $\hat{F}_{t,i}$ is unbiased, then the composite short term relative $\hat{F}_t$ will also be unbiased, and $E(\gamma_t) = 0$. In this case, the first term of equation 16 is zero. Also, the higher order terms of equation 16 will be central moments. On the other hand, if the sample short term relatives $\hat{F}_{t,i}$ are biased, the $\hat{F}_t$ will be biased, and $E(\gamma_t) \neq 0$. Then all terms of equation 16 will be larger, and rel-bias$(\hat{I}_T)$ may be quite large.

Possible sources of bias

It is important then that each $\hat{F}_{t,i}$ have a negligible bias for $\hat{I}_T$ to be a stable estimator of I_T. In the IPP application, there are three possible sources of bias that need to be watched:

1. The formal bias. The fixed-weight Laspeyres short-term-relative we wish to measure is

$$F_t = \frac{\sum_{i=1}^{N} W_{i,o}\, R_{i,t,o}}{\sum_{i=1}^{N} W_{i,o}\, R_{i,t-1,o}}$$

Each sample j provides an estimate of F_t. Suppose sample j's base quarter is t_j. (For example, on page 4 sample 2's base quarter is quarter 8.) The expectation of $\hat{F}_{t,j}$ at full-response is:

$$E(\hat{F}_{t,j}) \approx \frac{\sum_{i=1}^{N} W_{i,t_j}\, R_{i,t,t_j}}{\sum_{i=1}^{N} W_{i,t_j}\, R_{i,t-1,t_j}} = F_{t_j,t}$$

We are using an updated market basket: the market basket based on quarter t_j rather than quarter 0. As stated before, we are uncertain about this formal bias: part of the reason we are sampling every two years is that we wish to obtain data from new product areas and new markets. However, the composite estimator

$\sum_{j=1}^{3} w_j\, \hat{F}_{t,j}$ has expectation $\sum_{j=1}^{3} w_j\, F_{t_j,t}$

which is not F_t. The composite estimator is an unbiased measure of a weighted average of population short-term-relatives, each based on a slightly different market basket. This state of affairs was considered acceptable, so it will be assumed that $\sum_{i=1}^{3} w_i\, \hat{F}_{t,i}$ is unbiased from a formal viewpoint.

2. Non-response bias. The numerator and denominator of $\hat{F}_{t,i}$ are both Horvitz-Thompson estimators of the corresponding numerator and denominator of the population short term relative F_t (ignoring the formal bias). Therefore at full response they are unbiased estimates. Since the IPP is a voluntary program, many sample units do not respond. Therefore using the original probabilities of selection in the Horvitz-Thompson estimator will no longer guarantee unbiased estimates. It is important to adjust the weighting structure of the responding units based on a response model to restore unbiasedness as much as is feasible. The adjustment should be made within each sample i.

3. The ratio estimate bias. This final source of bias is the technical bias arising from the non-linearity of the estimates $\hat{F}_{t,i}$. This bias is small if the sample size for each sample estimate $\hat{F}_{t,i}$ is reasonable (see Cochran,pp. 160-162). If the sample size is very small, then in certain cases, this bias could become significant. One should not use this methodology with very small samples. (The extreme case would be a set of samples of size 1 from the frame: the bias would be large in this case). As stated in Cochran, p. 160, this bias can be ignored if each sample i has even a moderate sample size.

In conclusion, bias can be a serious problem in the estimate $\hat{I}_T$ if each $\hat{F}_{t,i}$ has any more than a very negligible bias. Therefore, one needs to assure oneself that $\hat{F}_{t,i}$ has a small bias. A reasonable sample size for each sample should remove the third source of bias, and the second source can be reduced by careful weight adjustment within each sample.

MSE$(\hat{I}_T)$ within unbiased $\hat{F}_t$'s

In this section, it will be assumed that the problems stated above are negligible and each $\hat{F}_{t,i}$ is in fact unbiased. If each $\hat{F}_{t,i}$ is unbiased, equation 14 gives:

$$\frac{\text{Bias}(\hat{I}_T)}{I_T} = \sum_{t=1}^{T} \sum_{\substack{s=1 \\ s \neq t}}^{T} \frac{\text{Cov}(\gamma_t,\, \gamma_s)}{F_t\, F_s} + \text{higher-order central moments} \qquad (18)$$

We can approximate the MSE of I_T also. By squaring equation 14 and taking expectations we obtain:

$$\frac{\mathrm{MSE}(\hat{I}_T)}{I_T^2} \approx \sum_{t=1}^{T} \frac{\mathrm{Var}\ (\gamma_t)}{F_t^2} + \sum_{t=1}^{T}\sum_{s=1}^{T} \frac{\mathrm{Cov}(\gamma_t,\ \gamma_s)}{F_t\ F_s}$$

$$+ \text{ higher-order central moments} \qquad \text{(eq 19)}$$

All terms of both equations 18 and 19 are of order 1/n. This means the squared bias decreases with order $1/n^2$ and the MSE decreases with order 1/n. Therefore, with large samples, bias will be a negligible component of overall MSE. The estimation methodology strives to set sample weights that minimize the first term of equation 19. If the second term is not negligible, however, then the estimation methodology may not be optimum. The following paragraphs examine the conditions under which the second term may be large. It will be shown that the second term could be large if there is a relationship between the autocovariances of the sample short term relative estimates $\hat{F}_{t,i}$ and the sample weights $w_{t,i}$ assigned to these estimates.

Each term of interest in equation 19 has factors of the form $\mathrm{Cov}(\gamma_t, \gamma_s)$, which is the same as $\mathrm{Cov}(\hat{F}_t, \hat{F}_s)$. $\hat{F}_t$ in turn is given by equation 10 (assuming the variance-covariance structure of the $\hat{F}_{t,i}$'s is known):

$$\hat{F}_t = (\underline{1}'\ \Sigma_{F,t}^{-1}\ \underline{1})\ \underline{1}'\ \Sigma_{F,t}^{-1}\ \underline{\hat{F}}_t = \underline{w}_t'\ \underline{\hat{F}}_t$$

The assumption of a known variance structure was made to make the $\underline{W}_t$'s fixed rather than random vectors. This will simplify the following work. We have then

$$\mathrm{Cov}(\hat{F}_t,\ \hat{F}_s) = \mathrm{Cov}(\underline{w}_t'\ \underline{\hat{F}}_t,\ \underline{w}_s'\ \underline{\hat{F}}_s) \qquad (20)$$

$\underline{\hat{F}}_t$ is a vector of $\hat{F}_{t,i}$'s, with each $\hat{F}_{t,i}$ having expectation F_t. Let $\varepsilon_{t,i}$ be the residual of $\hat{F}_{t,i}$, and $\underline{\varepsilon}_t$ be the vector of these residuals. Then $\underline{\hat{F}}_t = \underline{1}\ F_t + \underline{\varepsilon}_t$.

$\underline{W}_t$ and $\underline{W}_s$ are assumed fixed, so after rearrangement and expectation algebra equation 20 reduces to

$$\mathrm{Cov}(\hat{F}_t,\ \hat{F}_s) = \underline{w}_s'\ E(\underline{\varepsilon}_s\ \underline{\varepsilon}_t')\ \underline{w}_t \qquad (21)$$

Likewise the variance of each $\hat{F}_t$ can be rewritten as:

$$\mathrm{Var}(\ \gamma_t) = \mathrm{Var}(\ \hat{F}_t\ ) = \underline{w}_t'\ E(\ \underline{\varepsilon}_t\ \underline{\varepsilon}_t'\ )\ \underline{w}_t \qquad (22)$$

Using equations 21 and 22 we can rewrite equations 18 and 19:

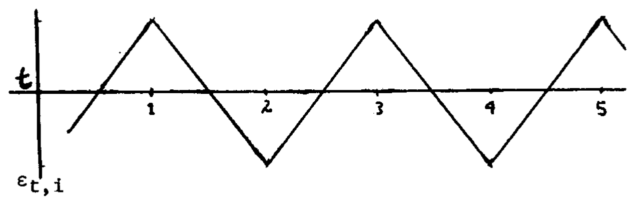
$$\frac{\mathrm{Bias}\ (\hat{I}_T)}{I_T} \approx \sum_{t=1}^{T}\ \sum_{\substack{s=1\\ s\neq t}}^{T} \frac{\underline{w}_t'\ E(\ \underline{\varepsilon}_t\ \underline{\varepsilon}_s'\ )\ \underline{w}_s}{F_t\ F_s} \qquad (23)$$

$$\frac{\mathrm{MSE}(\hat{I}_T)}{I_T^2} \approx \sum_{t=1}^{T} \frac{\underline{w}_t'\ E(\ \underline{\varepsilon}_t\ \underline{\varepsilon}_t'\ )\ \underline{w}_t}{F_t^2}$$

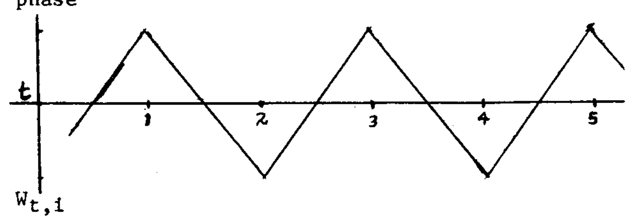
$$+ \sum_{t=1}^{T}\ \sum_{s=1}^{T} \frac{\underline{w}_t'\ E(\ \underline{\varepsilon}_t\ \underline{\varepsilon}_s'\ )\ \underline{w}_s}{F_t\ F_s} \qquad (24)$$

The nature of $\underline{w}_t'\ E(\ \underline{\varepsilon}_t\ \underline{\varepsilon}_s')\ \underline{w}_s$ might be made clearer by an example. Suppose for simplicity that the samples are independent, i.e. $\mathrm{Cov}(\varepsilon_{i,t},\ \varepsilon_{j,t}) = 0$ if $i \neq j$. Then the matrix $E(\underline{\varepsilon}_t\ \underline{\varepsilon}_s')$ in equations 23 and 24 will be diagonal.

Suppose there is a negative autocorrelation between $\varepsilon_{t,i}$ and $\varepsilon_{t,i+1}$ in sample i for all t. Then $\varepsilon_{t,i}$ might have this pattern:



If $W_{t,i}$ has the same pattern with the same phase



then $W_{t,i}$ will be high when $\varepsilon_{t,i}$ is positive and low when $\varepsilon_{t,i}$ is negative. A significant positive bias in $\hat{I}_T$ will result. If $W_{t,i}$ has the same frequency and opposite phase, then a negative bias in $\hat{I}_T$ will result. This is an intuitive explanation of the effect of $\underline{W}_t'\ E(\underline{\varepsilon}_t\ \underline{\varepsilon}_s')\ \underline{W}_s$ on Bias $(\hat{I}_T)$.

Suppose we are focusing on setting the sample weights $\underline{W}_T$ in quarter T. Previous quarters' $\underline{W}_t$'s have already been set and are fixed (this is reflective of the IPP application). Then from this viewpoint, any of the terms in equation 24 not involving $\underline{W}_T$ are fixed with regard to minimizing MSE $(\hat{I}_T)$ with respect to $\underline{W}_T$. Therefore, equation 24 can be rewritten as

$$\frac{MSE(\hat{I}_T)}{I_T^2} \approx Q(T-1) + \frac{\underline{W}_T' \, E(\underline{\varepsilon}_T \, \underline{\varepsilon}_T') \, \underline{W}_T}{F_T^2} + \sum_{s=1}^{T} \frac{\underline{W}_T' \, E(\underline{\varepsilon}_T \, \underline{\varepsilon}_s') \, \underline{W}_s}{F_T \, F_s} \quad (25)$$

Where Q(T-1) contains all terms of eq. 25 that do not include W_T.

The composite estimation methodology recommended essentially minimizes the second term, the variance term. In the IPP application, the third term can be assumed small if the sample size is moderate to large. Under these conditions, minimization of the second term above will be close to optimum.

If the sample size is small, the composite estimation methodology will not be optimum and $\underline{W}_T$ should be chosen to minimize the full formula. This requires estimates of $E(\underline{\varepsilon}_t \, \underline{\varepsilon}'_s)$, which will be difficult, especially when sample sizes are small. Therefore, one must require that sample sizes be at least moderate, and under this requirement $\underline{W}_T$ calculated under equation 12 will be near optimum.

IV. CONCLUSIONS

The composite estimator uses the various samples to generate estimates of the short term relatives in the present quarter. These estimates are the BLUE composite estimates of the true short term relative based on the vector of sample estimates. This methodology allows one to revise sample weights quarterly to gain maximum efficiency without causing breaks in the chain.

For this estimation methodology to be sound, each sample short term relative estimate should be unbiased. An adequate sample size is a necessary but not a sufficient condition for this to be true. There should also be concern about autocorrelative error patterns between short term relatives in the same sample. If these exist, then the sample weights in each quarter should be set with the autocovariances in mind. The BLUE will still minimize the variance of the short term relative estimate, but biases in the overall index may result. Adequate sample sizes should minimize the autocorrelative problem also, as research not presented in this paper has shown.

If the sample short term relatives are in fact unbiased and have white noise errors, then the short term chain methodology will give one unbiased $\hat{I}_T$'s with good variance properties.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the help of Rick Valliant, Marvin Kasper, Jim Himelein, Kim Zieschang, and other members of the BLS statistical staff in the development of this methodology and paper. Appreciation is also extended to Catreeda Lloyd for her typing work.

REFERENCES

[1] Banerjee, K., Cost of Living Index Numbers: Practice, Precision, and Theory, New York: Marcel Dekker, 1975.

[2] Cochran, R.S., "Multiple Frame Sample Surveys", 1964 Proceedings of the Social Statistics Section, American Statistical Association, 1964, pp. 245-249.

[3] Cochran, W.G., Sampling Techniques, New York: John Wiley & Sons, 1977.

[4] Hartley, H.O., "Multiple Frame Surveys", 1962 Proceedings of the Social Statistics Section, American Statistical Association,

[5] Kasper, M., "Import and Export Price Indexes Sample Design", BLS Statistical Notes, No. 16.

[6] Lund, R.E., "Estimation in Multiple Frame Surveys", 1968 Proceedings of the Social Statistics Section, American Statistical Association.

[7] Searle, S.R., Linear Models, New York: John Wiley & Sons, 1971.

[8] Theil, H., Principles of Econometrics, New York: John Wiley & Sons, 1971.