ESTIMATING LOW-VALUED EXPORTS FROM THE U.S.

Patrick J. Cantwell, Ryan M. Fescina, Melvin McCullough, U.S. Census Bureau[1]
U.S. Census Bureau, SRD, Washington, DC 20233-9100
patrick.j.cantwell@census.gov

KEY WORDS: truncated reporting, exemption level, missing data, composite coefficients, borrowing strength.

1. Introduction

Each month the U. S. Census Bureau receives documents describing all U.S. exports over $2,500. In an attempt to keep the burden of filing low, exporters are not required to report their "low-valued" exports (LVEs)--those valued at or under $2,500. Instead, we estimate the low-valued component for any country by multiplying its total of exports valued over $2,500 by a pre-determined factor. This factor, specific to the country of export, was determined in the late 1980's based on data collected up to that point in time. However, while export patterns--commodities, modes of transport, quantities, and values--have changed, the factors remain the same.

In this paper we propose a new method to estimate the low-valued component. Our goal is to present a method that (1) accurately represents what is currently being exported, and (2) can adapt over time with minimal review to reflect the inevitable changes in the patterns of exports.

We break the problem into three parts. First, we divide all exports for a country into several groups, hoping to combine those LVEs for which the ratios of low- to high-valued exports are fairly homogeneous. Next, we determine a set of observed exports on which to base the new estimate of LVEs. Each month some exporters provide information on all their shipments, even those below $2,500. Taken in the proper context, these data can be used to estimate the unknown LVEs. Finally, for many countries, there will be groups of exports (as defined later) for which there are few or no data on which to base an estimate. When this is the case, information can be drawn from elsewhere in the same block of countries.

The methods and analysis presented below ignore all exports to Canada. Because Canada tracks their imports so well, the Census Bureau uses Canadian import figures as estimates of our exports to Canada. On the other hand, some exports destined for countries outside North America are shipped first to Canada or Mexico, before continuing on their way. The final country of destination is the only one considered in U.S. export totals.

In Section 2, past and current procedures for reporting exports and estimating LVEs are reviewed. Section 3 covers how we determine effective variables for defining groups and how many groups to use. In Section 4, we address the question: Which observed data can best help one estimate the LVEs? In Section 5, a compositing procedure is proposed for combining information from other countries when data are scarce. Section 6 contains results of evaluations on the several proposed procedures.

2. Background: Prior and Current Methods of Estimating Low-Valued Exports

History of Estimating Low-Valued Exports

The Foreign Trade Division of the U.S. Census Bureau publishes the nation's official international trade statistics, including data on imports and exports. The Division began estimating LVEs in the 1960s. Starting then, to reduce processing costs and filer burden, the Bureau has set exemption levels; shippers have only been required to report complete export data for transactions valued greater than a specified exemption level. Exports valued below the exemption level have been estimated using country-specific factors. The LVEs for each country are produced by multiplying that country's factor by its previous month's export total above the cutoff.

From 1965 to 1989 the Census Bureau revised these factors every few years, or with each change in the exemption level. The updates reflected changing trade patterns. For a given country, this process of updating factors was done by examining the change--from the prior year to the current year--in the proportion of total trade accounted for in the range from the exemption level to twice the exemption level. The estimated change was multiplied by the existing factor to create the newly updated factor.

In January 1985, the exemption level was raised from $500 to $1,000, and the low-value factors were adjusted as a result. The existing factors for each country were used to estimate the total value of items valued $500 or below. Then a sample of data from 1984 was used to measure the total value between $501 and $1,000, and the total value greater than $1,000 for each country. The ratio of these two totals was added to the existing low-value

[1] This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

factor to create the new factor.

In order to improve the estimates, the Census Bureau began to calculate "world-area" factors in October 1989 for countries with very little trade by grouping data from neighboring countries. This change affected 33 countries. While the overall change to the total estimate was insignificant, several countries saw substantial changes in their estimates. Since this update in 1989, the Bureau has created low-value factors as needed for new countries using world-area factors. Existing factors have not received any further updates for fear that the periodic updates might actually worsen the estimates.

Development of the Automated Export System (AES)

Since the low-valued export factors were determined in the late 1980's, the system for collecting and processing export data has evolved. The Automated Export System (AES) is now the main avenue through which export shipment data are electronically filed to the Customs and Border Protection Service (Customs). A filer can be an exporter or an authorized forwarding agent.

AES provides exporters with an alternative to filing paper Shipper's Export Declarations (SEDs). Although initiated in 1995, electronic filing on AES has been possible for all ports and for all modes of transport since 1997. The number of exporters using AES has grown steadily. As of January 2003, 85.9% of all eligible export shipments were filed on AES. While using AES is currently optional, the Security Assistance Act of 2002 makes filing via AES mandatory for all shipments requiring an SED. Implementation for mandatory AES filing is planned for mid-2004.

Filing via AES provides many advantages and options that are not available with paper SEDs. Data collected on AES can be edited immediately to ensure that filers comply with current U.S. export reporting requirements, and to identify data reporting errors. The system informs filers of these errors and allows them to make corrections. This benefits filers, but also improves the Census export trade statistics.

Another feature of AES is providing options on how to file. Exporters can choose to file their data (1) at a summary level, where they must summarize all their shipment data valued $2501 or more by commodity number, or (2) at a detailed (non-summary) level, where each transaction--regardless of value--must be reported as an individual commodity line item. AES provides all these benefits, while simultaneously reducing costs associated with handling and keying paper documents, correcting errors, and duplicate reporting.

Table 1 provides a breakdown of exports in a typical month and how they are filed. One can see the dichotomy of concern to us: the detailed exports, for which LVEs are available; versus the summary exports (summary AES and paper), for which LVEs are not observed (shaded in the table). Currently, about 15% to 20% by volume of those using the AES provide detailed data, including information on their low-valued exports.

Our task is to estimate the total value of LVEs from summary reporters, those labeled "Not Observed" in Table 1. The actual value of LVEs from detailed reporters in any month is known and would be included as tabulated when estimating the overall level of LVEs. On the other hand, the estimated LVRs will be based on the totals of low- and high-valued exports from detail reporters, as these are the only figures for LVEs that are observed and available.

Table 1. Exports by Reporting Method and Size

Records in a Typical Month (excluding Canada)	Automated Export System (AES) 1,170,000 records observed		Non-AES (Paper) 180,000 records observed
	Detailed (Non-Summary)	**Summary**	**Summary** (only)
High-Valued Exports: > $2,500	230,000	710,000	180,000
Low-Valued Exports: ≤ $2,500	230,000	*Not Observed*	*Not Observed*

Notes:
1. Frequencies are based on an average of three months in 2002 and 2003, and rounded to the nearest 10,000.
2. Excluding Canada, non-AES (paper) data currently represent about 13% of exports, and 20% of total volume of exports. This portion should decrease to 0% over time.

3. Breaking Up All Exports into Groups

In applying low-value factors to the totals of high-valued exports (HVEs) from summary reporters within a country, we want the procedure to be robust to changes in export patterns, so that it automatically adjusts the estimate of LVEs. For example, suppose that the United States has been shipping heavily in grain and appliances to a certain country. The low- to high-value ratio (LVR) of exports for that country should reflect such an export history. But suppose that later the U.S. starts exporting computers and electronic equipment to that country. Without a change in the LVR, the new estimate of LVEs will be based on the old export pattern, implicitly assuming that the ratio of low- to high-valued exports tends to be the same across the different commodity groups.

Rather than make such a questionable assumption, we propose dividing all exports into groups j within countries so that the total of LVEs for country i can be estimated as the sum of the LVEs estimated within each group, as in equation (1):

$$\mathrm{LVE}_i = \sum_{j=1}^{G} \mathrm{LVE}_{ij} = \sum_{j=1}^{G} c_{ij}\, \mathrm{HVE}_{ij} \qquad \textbf{(1)}$$

Note: Because the LVEs are known for detailed filers, the expressions LVE_i, LVE_{ij}, and HVE_{ij} found in equation (1) and in the following sections refer to the totals of low- or high-valued exports *only* from summary reporters.

For the analyses in this paper, we examined export records, mainly data filed electronically via the AES. Referring to Table 1, AES filers have the option of filing all their records individually (that is, *detailed* or non-summary), or summarizing their data by commodity and shipment (*summary*). Those filers who elect to file on a detailed basis are expected to report all their records, regardless of value.

Variables That Influence the Low-to-High-Value Ratios

To determine the composition of the groups for computing LVRs, we investigated the effects of variables such as the mode of transport (MOT) and the commodity. In the following, each export's MOT takes one of three values: vessel, air, or over-land; over-land includes all exports that are not by vessel or air. MOT is defined to be the mode used as the export leaves the United States, even though the shipment may stop in more than one country and several modes may eventually be used. For example, if a shipment is brought to Canada by truck or rail before being sent by vessel to Ireland, the export's mode and destination are characterized as over-land to Ireland.

To describe types of commodities, we started with the two-digit codes or “chapters” of the Schedule B, the ten-digit hierarchical classification system used to collect and compile U.S. export statistics. Schedule B is based on the Harmonized Commodity Classification System (HS), which was developed by the world’s major exporting countries to facilitate international trade. The system offers a uniform structure for classifying products and documenting statistical information on customs tariffs and transportation characteristics. There are approximately 8,000 different commodity classifications in Schedule B.

We tested the hypotheses that MOT and chapter (HS) level influence the LVRs of detailed data by conducting analysis of variance (ANOVA) on the data. It should be noted that the usual assumptions of an ANOVA do not hold with these data. The distributions of LVRs need not approximate a normal or symmetric shape, and the issue of unequal variances might be a concern. Thus, we use the test not so much to derive a valid significance level from an F test, but more to indicate which characteristics appear to have greater influence on the LVRs.

The first ANOVA modeled LVRs as a function of country, MOT, and two-digit HS code, and included all two-way interactions. We limited the data to include 28 countries and 20 two-digit HS codes, each highly ranked in terms of total export value. The data came from detailed reports of exports from May 2002 to July 2002; the ratios were computed for each month. We found that the LVRs depend highly on country, MOT, and HS, as well as on all the two-way combinations.

We repeated many of the studies on data collected in October through December of 2002 to see how consistent the observed relationships are over time. In addition, we joined the data sets, attaching to each record an indicator of the time period, May-July or October-December. After repeating the earlier analyses, we tested for a difference in the LVRs from one time period to the next. Although country, MOT, and HS (as well as their two-way interactions) showed significant effects on the values of the LVRs, the difference over time was not statistically significant, lending support to the notion that the observed relationships are consistent over time.

Investigations into other variables established that some had a moderate effect on the LVRs, but not as strong as MOT and HS code (whether one-digit or two-digit HS code). For example, we considered a variation of the commodity codes called “section” numbers, used by analysts in the Bureau's Foreign Trade Division. Although the effect of section code on LVR was less significant than HS, it might work well as a proxy for HS.

How Many Groups?

How many groups should be used? With more groups, finer distinctions can be made within the set of exports, whether they be divided by commodity types or other characteristics. On the other hand, as the groups become small in number of shipments or total volume of exports, there are fewer data on which to base the LVR used in the estimate of the group's LVE, possibly leading to decreased stability in the estimate.

In addressing this issue, we mention the use of composite LVRs for countries that have too few exports in a group. As will be described in the Section 5, the composite factor combines these exports with data on exports *in the same group* but from other nearby countries. The result is a more stable estimated LVR based on the specific group characteristics. Our belief is that the LVR depends more on the type of export--for example, the type of commodity or method to ship it--than on the country of destination.

Based on these several considerations, we propose using 30 groups: three MOTs crossed with ten HS codes. To get an indication of how well this division works, we examined the estimates of LVEs for a number of countries. As will be seen in Section 6, dividing the exports into 30 groups appears to be a successful strategy.

4. Using AES Detailed Data to Make Inferences About Missing LVEs

As discussed above, our goal is to develop factors based on the LVR for the detailed filers, and then apply those factors to totals of HVEs for the summary reporters. These factors will take into account the exports' country of destination, the mode of transport, and the commodity. However, to support our use of the factors, we need to show that detailed filers are similar to, or representative of, the summary filers.

Without having access to the LVEs of summary filers, one must look at other data. Table 2 provides summary statistics--the two quartiles, the median, and the mean--of exports in several ranges according to the method of filing.

Table 2. Summary of Exports in Selected Ranges

	Range of Value of Exports				
	$0 - 2,500	$2,501- 5,000	$2,501- 10,000	$2,501- 20,000	> $2,501
	AES Detailed				
Q1	55	2,965	3,364	3,696	4,162
Median	185	3,509	4,579	5,611	7,562
Mean	397	3,594	5,080	7,231	32,360
Q3	522	4,189	6,470	9,664	18,223
	AES Summary				
Q1	*Not*	3,000	3,598	4,268	5,800
Median	*Avail-*	3,575	5,060	7,157	12,990
Mean	*able*	3,641	5,489	8,422	59,154
Q3		4,242	7,198	11,827	32,731
	Paper				
Q1	*Not*	3,000	3,500	3,968	4,772
Median	*Avail-*	3,550	4,849	6,300	9,669
Mean	*able*	3,637	5,308	7,787	46,584
Q3		4,224	6,885	10,654	23,829

Except for the first column of LVEs, the ranges are bounded below by $2,501, and above by $5,000, $10,000, $20,000, or ∞ (that is, no limit). (Results using an upper bound of $15,000 were analogous, and are omitted to save space.) Based on Table 2, it is easy to see that, as the interval's upper bound increases, the summary statistics are less alike across the three methods of filing. In fact, the statistics are very similar across the methods within the interval [$2,501, $5,000].

This result implies that, when addressing missing data for summary files based on detailed (observed) data, one may fare well by restricting the procedure to data in a range that does not greatly exceed that of the exemption level, $2,500. As always, there are other considerations. Defining HVEs with a wider range will provide a larger number of exports from detailed filers on which to base the estimated LVRs. In addition, it is possible that the exemption level, currently $2,500, might be raised at some time in the future. In Section 6, several upper limits for defining HVEs are examined according to their performance estimating LVEs. It will be seen there that the range ($2,500, $10,000] works well in our studies for estimating LVEs.

5. Deriving Compositing Factors c_{ij}

Because export traffic to some countries is heavy enough within the previously defined export groups, there are ample data to compute an LVR within that group. But for many countries and at least some export groups, the data are sparse enough to raise concern about the stability of the data from that country on which the estimated LVR is made. In an attempt to stabilize the resulting LVR, where fewer data are available from country i, information is drawn from elsewhere in the same block ($\mathcal{B}$) of countries as country i. We label the remainder of the block, excluding country i, as $b(i) = \mathcal{B} \sim i$, and define a composite coefficient for combining data:

$$c_{ij} \;=\; \lambda_{ij}\, f_{ij} \,+\, (1 - \lambda_{ij})\, f_{b(i),j}\,, \;\text{ where } 0 \le \lambda_{ij} \le 1 \quad \textbf{(2)}$$

The LVR, f_{ij}, can be defined in a number of ways. For the current discussion, we assume that there are M consecutive months of LVEs, x_{ijm}, and HVEs, y_{ijm}, $m = 1, 2, ..., M$. We estimate f_{ij} as the ratio of the totals of low- to high-valued exports:

$$f_{ij} \;=\; \frac{\sum_{m=1}^{M} x_{ijm}}{\sum_{m=1}^{M} y_{ijm}} \qquad \textbf{(3)}$$

Obviously, one could use alternative estimators, such as the average of the monthly ratios:

$$f'_{ij} \;=\; \frac{1}{M}\sum_{m=1}^{M} r_{ijm} \;=\; \frac{1}{M}\sum_{m=1}^{M} \frac{x_{ijm}}{y_{ijm}} \qquad \textbf{(4)}$$

For all countries in the block except i, the ratio $f_{b(i),j}$ can be defined analogously:

$$f_{b(i),j} = \frac{\sum_{k \in b(i)}^{B} \sum_{m=1}^{M} x_{kjm}}{\sum_{k \in b(i)}^{B} \sum_{m=1}^{M} y_{kjm}} \qquad \textbf{(5)}$$

or alternatively as an average of monthly ratios. As we will discuss later in this section, blocks can be composed of countries based on various criteria: geographic proximity, exhibiting similar LVRs for the export groups, or other characteristics.

First, we look for an appropriate value of λ_{ij}. We believe that, if sufficient data are available for the specified country and export group, then c_{ij} should be a "direct" LVR, that is, one determined completely by the data within country *i* and group *j*. This is equivalent to saying that λ_{ij} should be set to 1, leaving $c_{ij} = f_{ij}$. How many exports are sufficient is not an obvious matter. Based on our analyses of exports to various countries within the 30 export groups (see Section 6), we found little difference when using alternative "cutoff" values. We believe that \$1 million in HVEs, that is, exports in the range (\$2,500, \$10,000], would provide ample data with which to compute the LVR, and define the upper bound for compositing as such. Unless otherwise noted, this value is used in all analyses.

Turning to the more complex case where there are fewer exports, for a specific group *j* in country *i*, we would like λ_{ij} to satisfy some or all of the following:

(a) λ_{ij} should increase with the volume of exports to country *i*, or the proportion within the block, $\mathcal{B}$, with more data from country *i*, we want to weight the factor f_{ij} more heavily;

(b) λ_{ij} should decrease with the variability of the monthly LVR within country *i* over the months for which we have data; the more variable this monthly LVR is, the more we want to use data from the other countries in the block to enhance stability; and

(c) λ_{ij} should increase with the variability of f_{ij} among the block of countries; as the LVR factor f_{ij} varies more among the other countries in the block, we have less confidence in those data to provide a good estimate of LVR for country *i*.

Our plan then is to develop a coefficient, λ_{ij}, that accommodates these or similar criteria, justify its use statistically, and then test its performance. The coefficient should be properly scaled to reflect the importance we place on the criteria.

One can think of the observed export data for a given month as being the realization of a random mechanism from some unknown distribution. In this context, one can devise a solution by considering the general convex combination of two estimators for an unknown parameter θ:

$$\hat{\theta}_{combo} = \lambda\, \hat{\theta}_1 + (1-\lambda)\, \hat{\theta}_2\ ,\ 0 \le \lambda \le 1 \qquad \textbf{(6)}$$

As an estimator for θ, suppose we express the bias and variance of $\hat{\theta}_1$ as $bias_1$ and v_1^2, respectively, with analogous notation for estimator $\hat{\theta}_2$, and the covariance of the two estimators as v_{12}. Then it is easy to show that the value of λ that minimizes the mean squared error (MSE) of $\hat{\theta}_{combo}$ is

$$\lambda_{\min} = \frac{v_2^2 - v_{12} + bias_2^2 - bias_1 bias_2}{v_1^2 - 2v_{12} + v_2^2 + (bias_1 - bias_2)^2} \qquad \textbf{(7)}$$

For export data, equation (2) is a special case of (6). If the two estimators in (6) are independent, and the first is unbiased, then $\lambda_{\min}$ in (7) simplifies. For the first assumption, there is no correlation between the levels of exports for country *i* and the other countries in the block. As to the second, f_{ij} in (3) may well be an unbiased estimate of the true LVR for country *i*, depending on the volume of export data (x_{im}, y_{im}) collected and their distribution; f_{ij} is a ratio estimate and may exhibit some bias if the volume is too small.

At this point, as all ratios and coefficients refer to the specific export group *j*, we suppress the subscript *j* for simplicity. Inspecting equation (3), if we can assume that the pairs (x_{im}, y_{im}) of low- and high-valued exports are independent from one month to another, the variance of the ratio estimate f_i can be approximated as

$$\mathrm{Var}\,(f_i) \approx \frac{F_i^2}{M}\left[\frac{\sigma_{x,i}^2}{\mu_{x,i}^2} - 2\rho_i \frac{\sigma_{x,i}\,\sigma_{y,i}}{\mu_{x,i}\,\mu_{y,i}} + \frac{\sigma_{y,i}^2}{\mu_{y,i}^2}\right] \qquad \textbf{(8)}$$

where F_i represents the underlying LVR for country *i* (and group *j*) for any month *m*, the μ's represent the monthly averages of low- (*x*) or high-valued (*y*) exports, σ^2 their monthly variances, and ρ the correlation coefficient between x_{im} and y_{im}.

We can express the LVR for the entire block of countries (excluding country *i*) in (5) as

$$f_{b(i)} = \frac{\sum_{k \in b(i)}^{B} \sum_{m=1}^{M} x_{km}}{\sum_{k \in b(i)}^{B} \sum_{m=1}^{M} y_{km}} = \frac{\sum_{m=1}^{M} x_{b(i),m}}{\sum_{m=1}^{M} y_{b(i),m}} \qquad \textbf{(5}'\textbf{)}$$

where $x_{b(i),m}$ is the sum of the LVEs from all countries in the block, except *i*, for month *m*; $y_{b(i),m}$ is defined similarly for HVEs. Proceeding as before with f_i, the approximate variance of $f_{b(i)}$ can be expressed in a form analogous to that in (8), but including all appropriate data in the block except from country *i*.

To minimize the MSE of the composite factor c_{ij}, recalling (7), we select the compositing coefficient λ_{ij} (or λ_i here, still suppressing the export group index *j*) as

$$\lambda_i = \frac{\mathrm{Var}(f_{b(i)}) + bias^2(f_{b(i)})}{\mathrm{Var}(f_i) + \mathrm{Var}(f_{b(i)}) + bias^2(f_{b(i)})} \qquad \textbf{(9)}$$

The variances in (9) can be estimated by estimating the several components, as above. One can show that under reasonable assumptions, λ_i tends to increase with the volume of exports--consistent with (a) on the prior page. Similarly, λ_i tends to decrease as the volume of exports from the other countries in the block increases (unless the total volume of HVEs has surpassed the upper bound). On the other hand, as the variability over the months of the *level of exports* for country *i* increases, λ_i decreases, consistent with (b).

The bias of $f_{b(i)}$, $E(f_{b(i)}) - F_i$, is difficult to estimate when there are few exports to country *i* (in group *j*). However, a rough estimate is the value taken by $f_{(b(i)} - f_i$.

At this point, we insert an additional term related to the squared bias component in (9), to reflect our concern as expressed in (c). As the variability among the estimated LVRs, f_k, for the countries in the block $\mathcal{B}$ increases, we may want to decrease the level of compositing, that is, increase the value of λ_i. Thus, we insert the term into the numerator and denominator of (9),

$$\hat{\mathrm{V}}\mathrm{ar}(f_k) = \frac{\sum_{k=1}^{B} (f_k - \bar{f})^2}{B-1}, \qquad \textbf{(10)}$$

where $\bar{f}$ is the average of the f_k over the block $\mathcal{B}$. The procedure is evaluated with and without the term in (10).

The terms related to the squared bias of $f_{b(i)}$ are inserted into (9) with coefficients δ_1 and δ_2 attached:

$$\lambda_i = \frac{\mathrm{Var}(f_{b(i)}) + \delta_1 (f_{b(i)} - f_i)^2 + \delta_2 \hat{\mathrm{V}}\mathrm{ar}(f_k)}{\mathrm{Var}(f_i) + \mathrm{Var}(f_{b(i)}) + \delta_1 (f_{b(i)} - f_i)^2 + \delta_2 \hat{\mathrm{V}}\mathrm{ar}(f_k)} \qquad \textbf{(11)}$$

Various values of the δ's are tested in an attempt to minimize the MSE of the composite factor.

One might have thought that the minimizing coefficient λ depends instead (or as well) on the variability of the *low-value ratios*, $r_{im} = x_{im}/y_{im}$ and $r_{b(i),m} = x_{b(i),m}/y_{b(i),m}$. To investigate this further, we derived the minimizing λ_i defining the LVRs as the mean of the ratios over the *M* months, as in (4). For purposes of comparison, we call this application of f'_i and $f'_{b(i)}$ Method 2; the original derivation with f_i and $f_{b(i)}$ as in (3) and (5), respectively, is labeled Method 1. Following the development as before,

$$\mathrm{Var}(f_i') = \frac{\sigma^2_{r,i}}{M}, \qquad \textbf{(12)}$$

$$\mathrm{Var}(f_{b(i)}') = \frac{\sigma^2_{r,b(i)}}{M}, \qquad \textbf{(13)}$$

where $\sigma_{ri}^2 = \mathrm{Var}(r_{im})$ and $\sigma_{r,b(i)}^2 = \mathrm{Var}(r_{b(i),m})$.

The minimizing coefficient λ_i' is then defined as in (9). In this version, λ_i' depends on the variances of the ratios, r_{im} and $r_{b(i),m}$; it does not depend directly on the mean levels of the LVEs and HVEs, $\mu_{x,i}$, $\mu_{x,b(i)}$, $\mu_{y,i}$, and $\mu_{y,b(i)}$, respectively, although these values are related to the variances of the ratios. We continue with the computation as before, redefining components, such as $bias^2(f'_{b(i)})$, in terms of the f'_i and $f'_{b(i)}$.

Grouping Countries into Blocks

There are several ways one could combine countries into blocks for the purpose of borrowing strength, aimed at countries that realize fewer exports in some groups (mode of transport × commodity code). To derive an optimal statistical process, one might combine countries whose LVRs are most alike across the 30 groups. This could be addressed by applying one of various statistical clustering techniques.

However, we believe that it is advantageous in several ways to form blocks from countries that are geographically proximate, that is, generally on the same continent or part of the continent. Such a decision may facilitate the evaluation of estimates of low-valued exports using available data, as (a) many summaries of foreign trade data adhere to such geographic groupings, and (b) some trade experts restrict their realm of proficiency to specific geographic boundaries. We propose that countries be assigned to blocks defined by the following nine geographic regions: Mexico, Central and South America (including the West Indies), Europe (excluding Eastern Europe), Eastern Europe (including Russia), Africa, Southwest Asia, Southern Asia, the Pacific Rim and Islands, and Southeast Asia (including the South Pacific Islands).

If, within a group *j*, there are an insufficient number of HVEs across the block on which to base a satisfactory estimate of $f_{b(i)}$ or $f'_{b(i)}$, this ratio is determined from exports across all blocks in the world.

6. Results and Observations

To evaluate the proposed procedure with its various options, we constructed a set of actual export data. From the database of the Automated Export System for the months May 2002 through April 2003, we retained all exports from detail filers, and removed all exports from summary filers. Twelve months were used in an attempt to minimize seasonal effects on the analysis. By restricting the analysis to detailed-level exports, we have access to all such LVEs; we can derive estimates and compare them to the actual total of the LVEs. On the other hand, extending the inferences about the performance of the estimators to summary-level exporters requires the assumption--unproven--that the proportion of LVEs is the same for detailed and summary filers.

The data set was divided systematically into four subsets of equal size, with each record having the same chance of falling into any of the four subsets. The idea was to determine direct and composite factors from one subset, and then apply the procedures to the other three subsets. Using all four subsets to determine factors (and three subsets each time to evaluate the procedures), one obtains 12 quasi-independent applications on which to base tentative conclusions about the performance of the procedures. (When we evaluate the current procedure with its fixed LVR factor, there are only four different results, as it is applied to each of the four subsets.) This strategy allows one to get an idea of the variability of the procedures.

For the analyses portrayed in this paper, we focus on six countries in Europe. In Tables 3, 4, and 5, the first three are "large" countries, those to which we export a high volume of goods; the last three are smaller countries. The estimators have also been studied as applied to other countries--large and small, inside and outside Europe--often with similar results. In each of these tables, the number shown is the average absolute relative error, that is, the average over the applications of the absolute difference from the true total of LVE, divided by that true total, recorded as a percent.

Table 3 demonstrates the effect of dividing exports into different numbers of groups: 1, all exports together; 3, by mode of transport (MOT); 10, by commodity (HS) code; or 30, by MOT × HS. In almost all cases, by adding MOT for grouping, the error decreases or remains the same. The results are mixed when adding the commodity code, HS. With the results inconclusive for HS, we retain all 30 groups with the hope of ensuring a method that is more robust to unexpected export changes in the future.

Table 4 evaluates the proposed procedure under different upper bounds for the range of HVEs: $5,000, $10,000, and ∞ (no limit, as in the current procedure). Comparisons are also made with and without compositing for upper bounds of $5,000 and $10,000.

The results demonstrate that allowing a composite factor improves the technique's performance sharply for the large countries, and to a lesser extent for the smaller countries. When defining the upper bound for the range of HVEs, and restricted to those cases where compositing is used, $10,000 works better than $5,000 (and $20,000, not shown) among these countries.

Table 3. How Many Groups? Which Variables?

Average Absolute Relative Error[1]	Number of Groups of Exports: 1	3 MOT	10 HS Code	30 MOT × HS
U.K.	2.3%	2.0%	2.4%	2.1%
Germany	0.9%	0.7%	0.8%	0.8%
France	2.7%	2.6%	2.4%	2.0%
Denmark	15.1%	14.2%	21.3%	16.6%
Portugal	8.1%	6.1%	6.7%	6.6%
Greece	18.2%	13.4%	13.8%	14.5%

[1] Using Ratio Method 1, with HVEs defined as $2,501 - $10,000, and $\delta_1 = 0.5$ and $\delta_2 = 0$.

Table 4. Whether to Composite; Defining HVEs

Average Absolute Relative Error	No Compositing[1] HVEs Define as			Compositing[1,2] HVEs Defined as	
	$2,501 - $5,000	$2,500 - $10,000	> $2,500	$2,500 - $5,000	$2,500 - $10,000
U.K.	8.2%	8.2%	21.1%	2.7%	2.1%
Germany	6.9%	6.9%	14.5%	1.5%	0.8%
France	9.1%	9.1%	9.7%	3.6%	2.0%
Denmark	16.9%	18.5%	15.2%	19.1%	16.6%
Portugal	15.0%	7.7%	33.2%	11.7%	6.6%
Greece	18.7%	19.3%	22.7%	16.2%	14.5%

[1] Using Ratio Method 1 with 30 groups. [2] With $\delta_1 = 0.5$ and $\delta_2 = 0$.

Table 5 compares the current procedure to the proposed procedure. For the latter, we compare two ways of defining the ratio estimator, Methods 1 and 2, as described above in (3) and (4).

In Table 5, for each country the current procedure sees a large average relative error. Although not shown in the table, each estimate is greater than the true value on the file of detailed exports. It appears that the current factors, developed in the late 1980's based on export patterns at the time, no longer reflect well the current patterns of exports. However, it is unclear whether the current procedure also overestimates LVEs from exporters who file in a summary fashion in the AES or on paper.

Table 5. Current vs. Proposed Procedure; Method 1 vs. Method 2

Average Absolute Relative Error	Current Procedure	Proposed [1] Procedure	
		Ratio Method 1	Ratio Method 2
U.K.	898.5%	2.1%	4.9%
Germany	218.8%	0.8%	8.4%
France	328.5%	2.0%	6.3%
Denmark	1172.7%	16.6%	16.2%
Portugal	340.6%	6.6%	10.3%
Greece	776.9%	14.5%	28.3%

[1] With 30 groups, HVEs defined as \$2,501 - \$10,000, and $\delta_1 = 0.5$ and $\delta_2 = 0$.

From Table 5, one also sees that Ratio Method 1 generally works better than Method 2. For Denmark, Method 2 works slightly better. A review of the errors under Ratio Method 1 imply that it works quite well under these circumstances. For the larger countries, the average absolute relative error across the 12 evaluations is about a percent or two, and the errors themselves appear to be somewhat balanced around 0 (not shown).

Different Coefficients for δ_1, δ_2, and Other Parameters

Other analyses examined the new method under different values of the parameters. A cutoff for compositing was defined as the maximum value of the total of HVEs (over the 12 months in the database) for which one would composite within a country and group. We studied the estimators with cutoff values of \$0.3m, \$0.5m, \$0.7m, \$1.0m, and \$1.5m, and found very little difference, although the smaller values did not fare quite as well. Our recommendation is to use \$1 million as the cutoff.

In equation (11), δ_1 is the coefficient of $(f_{b(i)} - f_i)^2$. Although we had reservations about using this term, it appears to help the estimation; values in the range 0.5 to 1.0 work well. On the other hand, δ_2, the coefficient of $\hat{\mathrm{Var}}(f_k)$ in (11) appears to work best when $\delta_2 = 0$, suggesting that this term be deleted.

How Will the Proposed Procedure Work Over Time?

Will the procedure perform as well as export patterns change over coming months and years? Probably not quite as well. But, as the volume of exports changes from one commodity or mode of transport to another, the factors in the different groups are applied to smaller or larger totals of HVEs, keeping the procedure current. Over time, some deterioration will likely occur *within* groups. That is, for the *same* commodities and mode of transport in a group, the relationship between LVEs and HVEs may change gradually, yielding a different underlying LVR.

7. Continuing Work

Perhaps the most important work still to be done is an examination of past studies that tried to reconcile exports from the U.S. with imports to several specific countries. Some of these studies have suggested that the current procedure may be underestimating LVEs from the U.S. As this is not consistent with what was seen for detailed exports in the results of Section 6, several questions arise: Is the distribution of LVEs recorded by detailed filers like that shipped by summary filers? For some reason--perhaps due to the commodity or preferred mode of transport--are shippers of LVEs more likely to file in a summary manner than shippers of HVEs? Does the possible underreporting of exports differ between low- and high-valued exports?

A second issue is estimating the compositing coefficient, λ. As one can see from equations (8), (9), and (11), estimating λ requires one to estimate several variances and other terms, introducing additional error. Other approaches--perhaps depending more directly on the volume of shipments used to estimate the low-valued ratios--should be investigated to determine an effective coefficient λ.

Finally, we have studied estimates of LVEs for the six countries mentioned in Section 6 and several others. The next step is to examine estimates of LVEs for all other countries, as well as estimates of total exports from the U.S. For many countries to which we export only a limited amount of goods, the effectiveness of the compositing procedure might be a serious factor in evaluating the overall proposal for estimating LVEs.

ACKNOWLEDGMENTS

The authors thank Debra Coaxum for her assistance in this work, and Carma Hogue and Richard Moore, Jr. for providing helpful comments on the paper.