Application of the Bootstrap Method in the International Price Program

Patrick A. Bobbitt, Steven P. Paben, MoonJung Cho, Te-Ching Chen,
James A. Himelein Jr., and Lawrence R. Ernst
U.S. Bureau of Labor Statistics
2 Massachusetts Avenue NE, Washington DC 20212 U.S.A.
bobbitt_p@bls.gov

Abstract

The International Price Program (IPP) collects data on United States trade with foreign nations and publishes monthly indexes on the import and export prices of U.S. merchandise and services. The IPP employs a three stage PPS design in which establishments, then broad product categories traded within establishments, and finally items within a category, are selected. Certainty selections can occur in the first two stages.

We present three variations of the bootstrap rescaling method adapted to the IPP sample design: 1) sampling at the first stage, treating certainty units as probability units, 2) sampling that allows for certainties, and 3) a procedure that extends the previous method, by collapsing single item strata.

Finally, we compare the stability, bias and coverage rates of the three approaches by simulating 1000 samples of a simulated universe using the IPP sampling methodology.

KEY WORDS: Variance estimation; Bias estimation; Certainty sampling units; Collapsing

1. Introduction

The International Price Program collects data on the United States' trade with foreign nations and publishes monthly indexes on the changes in import and export prices for both goods and services. Recently a research group was chartered to assess the quality of the existing variance measures[14], and if possible, propose alternate methods. This group studied the use various estimation methods: the Bootstrap, the Jackknife, BRR, and two methods derived from applying the Taylor's series[2].

In this paper, we will present an overview of the IPP sample design, weight structure and index estimation. We will then present the three bootstrap methods adapted to the IPP sample design. Finally, we will apply the methods to seven representative strata published by the IPP.

The International Price Program of the Bureau of Labor Statistics (BLS) produces two of the major price statistics for the United States: the Import Price Indexes and the Export Price Indexes. The IPP, as the primary source of data on price change in the foreign trade sector of the U.S. economy, publishes index estimates of price change for internationally traded goods using three primary classification systems - Harmonized System (HS), Bureau of Economic Analysis End Use (BEA), and North American Industry Classification System (NAICS). IPP also publishes selected services indexes and goods indexes based upon the country or region of origin. This paper will only focus on the Import goods indexes that IPP publishes monthly.

The target universe of the import indexes consists of all goods purchased from abroad by U.S. residents. Ideally, the total breadth of U.S. trade in goods in the private sector would be represented in the universe. Items for which it is difficult to obtain consistent time span for comparable products, however, such as works of art, are excluded. Products that may be purchased on the open market for military use are included, but goods exclusively for military use are excluded.

2. Sampling in the International Price Program

The import merchandise sampling frame used by the IPP is obtained from the U.S. Customs and Border Protection (USCBP). This frame contains information about all import transactions that were filed with the USCBP during the reference year. The frame information available for each transaction includes a company identifier (usually the Employer Identification Number), the detailed product category (Harmonized Tariff number) of the goods that are being shipped and the corresponding dollar value of the shipped goods.

IPP divides the import universe into two halves referred to as panels. One import panel is sampled each year and sent to the field offices for collection, so the universe is fully re-sampled every two years. The sampled products are priced for approximately five years until the items are replaced by a newly drawn sample from the same panel. As a result, each published index is based upon the price changes of items from up to three different samples .

Each panel is sampled using a three stage sample design. The first stage selects establishments independently proportional to size (dollar value) within each broad product category (stratum) identified within the harmonized classification system (HS).

The second stage selects detailed product categories (classification groups) within each establishment - stratum using a systematic probability proportional to size

(PPS) design. The measure of size is the relative dollar value adjusted to ensure adequate coverage for all three published strata across all classification systems, and known non-response factors (total company burden and frequency of trade within each classification group). Each establishment - classification group (or sampling group) can be sampled multiple times and the number of times each sampling group is selected is then referred to as the number of quotes requested.

In the third and final stage, the field economist, with the cooperation of the company respondent, performs the selection of the actual items for use in the IPP indexes. Using the entry level classification groups selected in the second stage, a list of items is provided by the respondent to the field economist. Using a process called disaggregation, items are selected from this list with replacement to satisfy the number of item quotes requested for each entry level classification group.

3. Index Estimation

IPP uses the items that are initiated and re-priced every month to compute its price indexes. These indexes are calculated using a modified Laspeyres index formula.

The modification used by the IPP differs from the conventional Laspeyres index by using a chained index instead of a fixed-base index. Chaining involves multiplying an index (or long term relative) by a short term relative (STR). This is useful since the product mix available for calculating price indexes can differ over time.

These two methods produce identical results as long as the market basket of items does not change over time and each item provides a usable price in every period. In fact, due to non-response, the mix of items used in the index from one period to the next is often different. The benefits of chaining over a fixed base index include a better reflection of changing economic conditions, technological progress, and spending patterns, and a suitable means for handling items that are not traded every calculation month.

Below is the derivation of the modified fixed quantity Laspeyres formula used in the IPP.

$$\begin{aligned}
LTR_t &= \left(\frac{\sum p_{i,t} q_{i,0}}{\sum p_{i,0} q_{i,0}}\right)(100) \\
&= \left(\frac{\sum p_{i,0} q_{i,0} \left(\frac{p_{i,t}}{p_{i,0}}\right)}{\sum p_{i,0} q_{i,0}}\right)(100) \\
&= \left(\frac{\sum w_{i,0} r_{i,t}}{\sum w_{i,0}}\right)(100) \\
&= \left(\frac{\sum w_{i,0} r_{i,t}}{\sum w_{i,0} r_{i,t-1}}\right)\left(\frac{\sum w_{i,0} r_{i,t-1}}{\sum w_{i,0}}\right)(100) \\
&= \left(\frac{\sum w_{i,0} r_{i,t}}{\sum w_{i,0} r_{i,t-1}}\right)\left(_{t-1}\right) \\
&= (STR_t)\,(LTR_{t-1})
\end{aligned}$$

where,

$$\begin{aligned}
p_{i,t} &= \text{price of item } i \text{ at time } t \\
q_{i,0} &= \text{quantity of item } i \text{ in base period } 0 \\
w_{i,0} &= (p_{i,0})\,(q_{i,0}), \text{ the total revenue in base period } 0 \\
r_{i,t} &= \frac{p_{i,t}}{p_{i,0}}, \text{ or the long term relative of item } i \text{ at time } t \\
LTR_t &= \text{long-term relative of a collection of items at time } i \\
STR_t &= \left(\frac{\sum w_{i,0} r_{i,t}}{\sum w_{i,0} r_{i,t-1}}\right)
\end{aligned}$$

For each classification system, IPP calculates its estimates of price change using an index aggregation structure (i.e. aggregation tree) with the following form:
Upper Level Strata
Lower Level Strata
Classification Groups
Weight Groups (i.e. Company-Index Classification Group)
Items

As mentioned previously, at any given time, the IPP has up to three samples of items being used to calculate each stratum's index estimate . Currently the IPP combines the data from these samples by 'pooling' the individual estimates.

Pooling refers to combining items from multiple samples at the lowest level of the index aggregation tree. These combined sample groups are referred to as a weight group. Different sampling groups can be selected for the same weight group across different samples, so it is possible that multiple items from different sampling groups can be used to calculate a single weight group index. This weight group level aggregation is done primarily so the Industry Analysts within IPP can perform analyses on the index information across samples.

4. Variance Project

The variance project was chartered in the IPP to find a variance estimation algorithm that would be useful given the unique aspects of both the IPP sample design and the modified Laspeyres index employed in the program. This project would analyze several different variance methods for both their precision and their bias through simulation. To achieve this, we drew one thousand complete samples from a the sampling frame that included all import transactions from July 2002, to June 2003. We then simulated the response data at the item level for these 1000 draws for a period of three years[2].

In this paper, we will inspect only the variance estimation of the short-term relative, specifically, the short term relative for the first month in a 'chain'. The reason for this is at the first month the LTR is the same as the STR, and due to the 'memory' effect of the LTR over time errors in estimating variance get compounded over time, complicating the analysis.

5. Bootstrap Methods

5.1 Literature Review

The IPP evaluated different variance estimation methods such as Taylor Series Linearization, bootstrap, jackknife, balanced repeated replication (BRR) for their applicability to the IPP. In this paper, we present three variations of the bootstrap rescaling method which were adapted to the IPP sample design in detail.

The bootstrap method for the iid case has been extensively studied since Efron proposed his bootstrap method in 1979, and considered as the most flexible method among well known resampling methods. The original bootstrap method was then modified to handle complex issues in survey sampling, and results were extended to cases such as stratified multistage designs [10].

Rao and Wu [11] provided an extension to stratified multistage designs but covered only smooth statistics. The main technique which was used to apply the bootstrap method to complex survey data was scaling. The estimate of each resampled cluster was properly scaled so that the resulting variance estimator reduced to the standard unbiased variance estimator in the linear case.

Sitter [13] also explored the extensions of the bootstrap to complex survey data and proposed a mirror-matched bootstrap method for a variety of complex survey designs. Sitter mentioned in his study that it was difficult to compare the performances of his proposed method with those in Rao and Wu [11]'s rescaling method either theoretically or via simulation.

Later, Rao et al [12] extended the result of Rao and Wu [11] to non-smooth statistics such as the median by making the scale adjustment on the survey weights rather than on the sampled values directly. Although this method is known to overestimate the true variance, it has an advantage that it does bootstrap-sampling only at primary sampling unit level.

5.2 Bootstrap Rescaling Method (BRM)

This method resamples only items selected in the first stage of sampling. Despite its simplicity, it is still unbiased, and consistent in the linear case. This method also works well when estimating 'non-smooth' functions such as percentiles.

To capture the variability induced by items originally sampled in the later stages, it adjusts the final stage's weight to take into account the 'other' variation that would have been captured when selected in a particular bootstrap sample. In IPP's case, this means that for each stratum, we will select the establishment via simple random sampling with replacement (SRSWR), then adjust item weight using the following formula:

$$w_{heci}^{b} = w_{heci}\left[\left\{1-\sqrt{\frac{n_n^b}{n_h-1}}\right\}+\sqrt{\frac{n_n^b}{n_h-1}}\left(\frac{n_h}{n_h^b}\right)m_{he}^{b}\right]$$

Where:

w_{heci}^b = The new b^{th} bootstrap weight for item i within Classif-Group c, traded by establishment e, in stratum h

w_{heci} = The original item weight for i, e, c, h defined above.

n_h^b = The number of items selected in the b^{th} bootstrap sample in stratum h.

n_h = The number of establishments sampled within stratum h

m_{he}^b = The number of times establishment e was selected in stratum h in bootstrap sample b.

In our case, we are going to select $n_h^b = n_h - 1$ items in a bootstrap sample, reducing the bootstrap weight defined above to:

$$\begin{aligned} w_{heci}^{b} &= w_{heci}\left(\tfrac{n_h}{n_h^b}\right)m_{he}^{b} \\ &= w_{heci}\left(\tfrac{n_h}{n_h-1}\right)m_{he}^{b} \end{aligned}$$

5.3 BRM Algorithm

1. Draw $n_h^b = n_h - 1$ items with replacement from the n_h establishments sampled in stratum h. Let m_{he}^b be the number of times establishment e was selected in stratum, h for bootstrap sample, b.

2. Define the bootstrap weights, thus:

   $$w_{heci}^{b} = w_{heci}\left(\frac{n_h}{n_h-1}\right)m_{he}^{b}$$

   Where the terms are defined as in the above definition of the bootstrap item weight

3. Compute, $\hat{\theta}_{hb}$, the price STR index for stratum h using the bootstrap item weights found in bootstrap sample b.

4. Iterate steps 1 through 3 for $b = 1, ..., 150$

5. Compute the bootstrap estimator:

   $$v_h = \frac{1}{150}\sum_{b=1}^{150}\left(\hat{\theta}_{hb} - \hat{\theta}_h\right)^2$$

   where $\hat{\theta}_h$ is the STR index for stratum h computed using the original sample

5.4 Bootstrap variance estimator using Rescaling Method with Certainties (BRMC)

This method is an extension of the BRM method seeking to address the issue of certainty selections used the IPP sampling methodology by obtaining bootstrap samples from the first stage of *variability*. We achieved the adjustment for certainties by obtaining the bootstrap sample from the first stage in which there was no certainty selection.

Prior to applying the (BRMC) algorithm, we must first partition S_h, the set of all items sampled from sampling stratum, h, into three groups:

$$h_1 = \left\{ \begin{array}{l} \text{The set of items in } S_h \text{ the establish-} \\ \text{ment was selected with a probability} \\ \text{less than 1.} \end{array} \right\}$$

$$h_2 = \left\{ \begin{array}{l} \text{The set of items in } S_h \text{ in which the} \\ \text{associated establishment was a} \\ \text{certainty selection, but the CG was} \\ \text{selected with a probability less than 1.} \end{array} \right\}$$

$$h_3 = \left\{ \begin{array}{l} \text{The set of items in } S_h \text{ in which both} \\ \text{of the associated establishment and} \\ \text{CG were certainty selections.} \end{array} \right\}$$

Using this definition, we compute a bootstrap item weight, $w^b_{h_j i}$ from the existing item weight when $n_{h_j} > 1$:

$$w^b_{h_j i} = w_{h_j i} \left[\left[\left\{ 1 - \sqrt{\frac{n^b_{h_j}}{n_{h_j} - 1}} \right\} + \sqrt{\frac{n^b_{h_j}}{n_{h_j} - 1}} \left(\frac{n_{h_j}}{n^b_{h_j}} \right) m^b_{h_j} \right] \right]$$

Where:

$w^b_{h_j i}$ = The new b^{th} bootstrap weight for item i, within sampling stratum partition h_j

$w_{h_j i}$ = The existing item weight for item i within sampling stratum partition h_j

$n^b_{h_j}$ = The number of entities selected for bootstrap sample, b in sampling stratum partition h_j. An entity is defined by the partition, j and will be detailed below.

n_{h_j} = The number of entities selected within sampling stratum partition h_j The entities depend upon the partition, j and will be defined below.

$m^b_{h_j}$ = The number of times the entity was selected in sampling stratum partition, h_j in bootstrap sample b.

$$\text{entity} = \left\{ \begin{array}{ll} \text{establishments} & j = 1 \\ \text{CG} & j = 2 \\ \text{items} & j = 3 \end{array} \right\}$$

In our case, we are going to select:

$$n^b_{h_j} = \left\{ \begin{array}{ll} n_{h_j} - 1 & n_{h_j} > 1 \\ 1 & n_{h_j} = 1 \end{array} \right\}$$

entities in bootstrap sample, b, reducing the bootstrap weight defined above to:

$$w^b_{h_j i} = \left\{ \begin{array}{ll} w_{h_j i} \left(\frac{n_{h_j}}{n_{h_j} - 1} \right) m^b_{h_j} & n_{h_j} > 1 \\ w_{h_j i} & n_{h_j} = 1 \end{array} \right\}$$

In the case where $n_{h_j} = 1$, we are assuming that such cases are likely absolute certainties that would have never produced any sampling variability. This is because if there is (for example) a company in which there is only a single CG beneath a certainty Establishment, it's most likely the only CG would have ever acquired.

5.5 BRMC Algorithm

1. Draw $n^b_{h_j}$ entities.

   case 1: $n_{h_j} > 1$ Select the entities using a simple random sample with replacement from the n_{h_j} entities in the sampling stratum partition h_j. Let $m^b_{h_j}$ be the number of times entity was selected in sampling stratum partition, h_j for bootstrap sample, b.

   case 2: $n_{h_j} = 1$ Select the single item.

2. Define the bootstrap weights, thus:

   $$w^b_{h_j i} = \left\{ \begin{array}{ll} w_{h_j i} \left(\frac{n_{h_j}}{n_{h_j} - 1} \right) m^b_{h_j} & n_{h_j} > 1 \\ w_{h_j i} & n_{h_j} = 1 \end{array} \right\}$$

   Where the terms are defined above.

3. Compute, $\hat{\theta}_{hb}$, the price STR index for published stratum h using the items from ALL sampling stratum partitions and the bootstrap item weights found in bootstrap sample b.

4. Iterate steps 1 through 3 for $b = 1, .., 150$

5. Compute the bootstrap estimator:

   $$v_h = \frac{1}{150} \sum_{b=1}^{150} \left(\hat{\theta}_{hb} - \hat{\theta}_h \right)^2$$

   where $\hat{\theta}_h$ is the STR index for published stratum s computed using the original sample.

5.6 Bootstrap variance estimator using Rescaling Method with Certainties and Collapsing (BRMCC)

This method extends the BRMC outlined above by including a simple collapsing algorithm for those variance strata that have only one entity available for selection. In the previous section, we chose to presume that in most cases single entity variance strata were for all practical purposes, certainties. This section treats these single cases as potential areas of sampling variability and will collapse to form a pool on which to sample.

This collapsing will be achieved by creating $\frac{n_{h_j}}{2}$ new variance strata, where n_{h_j} is the number of single entity variance strata in the sampling strata partition h_j. Each of these new variance strata will contain at least two entities, with possibly one containing three. Given the nature of the sample design we are likely to encounter this in the cases when $j = 2$ or $j = 3$. Example scenarios of this follow:

1. **j=2**: We find that in sampling stratum 052 we find eleven certainty establishments that have only sampled Classif Group. We will form five new variance strata four with two certainty establishments, one with three certainty establishments.

2. **j=3**: We find that in sampling stratum 052 we find 6 certainty establishments that also have certainty CGs with only one item. We will form three new variance strata each with two certainty estab||certainty CGs.

For each of these newly formed variance strata we will form them by pairing the $X_{[i]}, X_{[i+1]}$ where $X_{[i]}$ is the i^{th} order statistic of the dollar value for the entity.

5.7 BRMCC Algorithm

1. Identify the single entity variance strata

   This will be the j for which $n_{h_j} = 1$.

2. Sort the group identified in 1. above by the entity dollar value

3. identify the new variance strata as

$$h \| j \| i$$

   where h is the sampling stratum, j designates the partitioned group identified in section 2.1, and i is the greater rank of the entity dollar value for the pair. Example: Suppose that for sampling strata 52 we have six single entity variance strata in the second group (i.e. certainty establishments, but non-certainty CGs). both the top two establishment dollar value amounts would be coded 5221, the third and fourth largest would be coded 5222, the fifth and sixth largest would be coded 5223. In the case of an odd number of single entity variance strata for a given sampling stratum, the smallest group would contain three.

4. select bootstrap sample

   this is done as in the previous bootstrap example, only using the newly created variance strata

5. define the bootstrap weight

   again, as above

6. Compute $\hat{\theta}_{hb}$, the price STR index again as above.

7. iterate

8. compute bootstrap estimator as above

6. Analysis

6.1 Overview

Table 1 presents three two-digit strata that were used in our analysis. The first two, P07 and P87 were included because they were historically volatile, the third, P90 was randomly selected. This analysis will include a brief comparison of the performance of the three methods outlined to estimate the first month of a STR chain for the above strata using bias, stability, and coverage rate as criteria. For a more detailed study see (Chen, 2007).

We are limiting our analysis to only the first month of the STR chain. This study provides a more detailed analysis for individual strata than was provided in Chen's study.

Strata	Description	Reason included
P07	Edible vegetables, roots, and tubers	Historically Volatile
P87	Motor vehicles and their parts	Historically Volatile
P95	Toys, games and sports equipment; parts and accessories	Coverage close to BRR
P42	Articles of leather; travel goods, bags, etc.	Randomly Selected
P61	Articles of apparel and clothing accessories, knitted or crocheted	Randomly Selected
P90	Optical, photographic, measuring and medical instruments	Randomly Selected
P94	Furniture, stuffed furnishings; lamps, lighting fittings	Randomly Selected

Table 1: Studied strata descriptions

6.2 Comparison

For each of the methods we ran 150 iterations. For each method we computed the average Bias, Stability, and coverage the following way:

For each two-digit HS stratum k, let us define y_i be the full vector of entire sample i where $i = 1, \ldots, 1000$, $\hat{\theta}_{ki} = \hat{\theta}_k(y_i)$ and define

$$\bar{\hat{\theta}}_{k\cdot} = \frac{1}{1000} \sum_{i}^{1000} \hat{\theta}_{ki} \tag{1}$$

$$\tilde{V}_{k\cdot} = \frac{1}{1000-1} \sum_{i}^{1000} \left(\hat{\theta}_{ki} - \bar{\hat{\theta}}_{k\cdot} \right)^2 \quad \text{and} \quad \tilde{\sigma}_{k\cdot} = \sqrt{\tilde{V}_{k\cdot}}$$

As we do not have a true variance, for each of the two-digit HS stratum k, we use $\tilde{V}_{k\cdot}$ and $\tilde{\sigma}_{k\cdot}$ as our sampling variance and standard deviation.

Let $\hat{\sigma}_{mki}$ be the standard error estimator of a two-digit HS stratum k of sample i for the variance estimation method m. The relative bias of an interested variance estimation method is calculated as

$$\text{Relative Bias} = \frac{\left(\frac{1}{1000}\sum_i^{1000}\hat{\sigma}_{mki}\right) - \tilde{\sigma}_{k.}}{\tilde{\sigma}_{k.}} \times 100\%$$

and the stability is

$$\sigma(\hat{V}_{mki}) = \sqrt{\frac{1}{1000-1}\sum_i^{1000}\left(\hat{V}_{mki} - \bar{\hat{V}}_{mk.}\right)^2} \times 100\%$$

where $\bar{\hat{V}}_{mk.}$ is the average of the 1000 variance estimations for the method.

We formed 95% confidence limits using:

$$\begin{aligned} \hat{\theta}_{kiL} &= \hat{\theta}_{ki} - t_{0.975,\mu_{mki}/2}\hat{\sigma}_{mki} \\ \hat{\theta}_{kiU} &= \hat{\theta}_{ki} + t_{0.975,\mu_{mki}/2}\hat{\sigma}_{mki} \end{aligned}$$

where,

$\hat{\theta}_{ki} =$ the price relative for stratum k in sample, i

$\mu_{mki} =$ The degrees of freedom for the two digit price relative

From this, we form the coverage rate thus:

$$\hat{c} = \frac{1}{1000}\sum_i^{1000} I\left\{\bar{\hat{\theta}}_{k.} \in \left(\hat{\theta}_{kiL}, \hat{\theta}_{kiU}\right)\right\}.$$

where,

$$I = \left\{ \begin{array}{ll} 1 & \text{, if } \bar{\hat{\theta}}_{k.} \in \left(\hat{\theta}_{kiL}, \hat{\theta}_{kiU}\right) \\ 0 & \quad otherwise \end{array} \right\}$$

$\bar{\hat{\theta}}_{k.} =$ is the average of 1000 index estimates used as population ("true") index.

We are using the relative biases for accuracy and percentage of the stability for the variation because the biases and stability are all very small. Note that by the way we have defined stability, a lower number means greater stability, while a higher number means less stability.

6.2.1 Relative Bias

Table 2 displays the relative bias of each method as compared to the variance observed in the 1000 samples. In this table, there seems to be little difference in the relative bias between the BRM and BRMC methods. In both cases both methods have a negative bias. With the exception of strata P07, P95, P94 the BRMCC method moved the relative bias closer to zero.

For each method, the sign and magnitude of the bias seems to be independent of the size of the estimated population variance, $\tilde{\sigma}_{k.}$.

Strata	BRM	BRMC	BRMCC	$\tilde{\sigma}_{h.}$
P07	3.94%	4.37%	4.68%	0.0502
P42	-14.91%	-14.87%	-13.68%	0.0015
P61	-11.78%	-15.78%	-5.22%	0.0016
P87	-6.39%	-6.98%	-0.99%	0.0056
P90	-4.02%	-6.44%	0.08%	0.0018
P94	-0.31%	-0.46%	2.58%	0.0019
P95	16.98%	13.78%	18.75%	0.0019

Table 2: Rel. Bias Results for 1000 samples

6.2.2 Stability

Table 3 displays the stability (or variance of the variance) of each variance method. From this table, it is clear that the stability for all methods is roughly the same.

Strata	BRM	BRMC	BRMCC
P07	0.1478%	0.1464%	0.1509%
P42	0.0002%	0.0002%	0.0002%
P61	0.0002%	0.0002%	0.0003%
P87	0.0048%	0.0047%	0.0056%
P90	0.0003%	0.0002%	0.0003%
P94	0.0003%	0.0003%	0.0004%
P95	0.0016%	0.0005%	0.0005%

Table 3: Stability Results for 1000 samples

6.2.3 Coverage Rate

Table 4 displays the coverage rates for the 1000 samples for each of the three methods presented. From this we can see that with the exception of P61 and P90 the BRMCC method provided coverage rates closer to the expected 95% for all strata, and the BRMC method offered no improvement to the BRM method in all but two cases, P07 and P95, and these were so small as to be negligible.

Strata	BRM	BRMC	BRMCC
P07	0.944	0.948	0.95
P42	0.91	0.92	0.922
P61	0.921	0.886	0.919
P87	0.899	0.892	0.906
P90	0.955	0.945	0.965
P94	0.908	0.904	0.92
P95	0.922	0.926	0.931

Table 4: Coverage results for 1000 samples

6.3 Discussion

Since IPP has a rather complete frame provided from Customs, we could compute the sampling fraction for the

BRM and the BRMC methods. For the BRMCC method, we did not have actual item data available from the frame, making such a calculation impossible. Table 5 shows a summary of all sampling strata for the BRMC method. The h_1 column presents the number of items sampled, PSUs, and averages in the sample and frame when the variance PSU is defined to be the establishment. The h_2 column presents the number of items sampled, PSUs and averages in the sample and frame when the variance PSU is defined to be a classif group within a certainty establishment. A rough estimate of the sampling fraction can be seen by dividing the average number of elements to select per variance strata for the frame and sample we find that for probability establishments we have a ratio of 12.4% and for certainty establishments we have about 14.4%.

Type	$\mathbf{h_1}$	$\mathbf{h_2}$	**Total**
Total number Sampled	2299	1091	3390
Num. Variance PSU	72	435	507
Ave Elements in VPSU	31.93	2.51	6.69
Total number in Frame	18,512	7,519	26,031
Num. Variance PSU	72	433	505
Ave Elements in VPSU	257.11	17.36	51.55

Table 5: Average elements in variance strata

6.3.1 Impact Sampling Methodology on Variance

Table 6 shows the number of variance PSUs available to be selected at each variance stratum, within the two-digit stratum. This table further breaks out the number selected in each of the three possible selection classes for the BRMCC method.

From this table we see that P07 has no certainty variance PSUs. This provides evidence as to why there was negligible impact to both the relative bias and the coverage rates. Since this strata also had the least number of variance PSUs available to sample, which may explain why this strata had the largest stability results.

Note too, that P87 and P90 have the largest proportions of their variance PSUs residing in the h_3 category. That is, most of their VPSUs are items that were selected in the third stage after certainty selections in the first two stages. This provides some evidence as to why there was a clear impact on both the relative bias and coverage rates when using the BRMCC method.

The observation that the proportion of items having certainties at later stages in the sampling methodology are better suited using the BRMC, BRMCC methods is further substantiated by observing that P61 has a proportion of its variance PSUs somewhere between P07s and P87/P90 and its relative bias/coverage rates are similarly affected.

stratum	h_1	h_2	h_3	total
P07	20	0	0	20
P42	45	3	14	62
P61	165	57	75	297
P87	227	104	815	1146
P90	171	26	200	397
P94	164	17	105	286
P95	114	18	86	218
other	1393	209	989	2591
n_h	2299	434	2284	5017

Table 6: Number of PSUs per strata by type

6.4 Conclusions

There is some evidence to suggest that sampling from the first level of variability in the sampling methodology as opposed to the first stage of sampling has a positive effect on estimating variance in the IPP. The use of the BRMC method, however seems to be less effective than the more general BRMCC method. This is likely due to the high proportion of 'single item' certainty CGs that were set to certainties in the BRMC method as opposed to the collapsing scheme used in the BRMCC method.

6.5 Future Work

Try adapting the BRM derived by Rao et. al.[12] for the single stage without replacement sampling design as opposed to the *with* replacement version we adapted here.

In our study we did not take into account the impact of imputation, while we did simulate non-response to sampling based upon models developed in the IPP, we presumed that respondents who agreed to participate would do so every month, thus negating the need to impute. Future work should include studying the impact of imputation.

In the IPP, our design was developed to support multiple aggregation structures. Since we only sampled one panel, it was 'difficult' to address the affect of using an alternate aggregation structure would have on the variance of indexes produced for these structures. Future work will address the impact of alternate aggregation structures.

Similarly to the 'alternate' classification system question, our simulation was applied only to a single sample. In production we actually 'pool' data from multiple samples. Future work will address the impact of pooling multiple samples.

Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.

References

[1] Bobbitt, P., Cho, M.J. and Eddy, R. M. "Weighting Scheme Comparison in the International Price Program." 2005 Proceedings of the American Statistical Association. *Government Statistics Section* [CD-ROM], 1006-1014 (2005).

[2] Chen, Te-Ching; Bobbitt, Patrick; Himelein, James A. Jr.; Paben, Steven P.; Cho, Moon Jung; Ernst, Lawrence "Variance Estimations for International Price Program Indexes." 2007 Proceedings of the American Statistical Association

[3] Cho, M. J., Chen, T-C, Bobbitt, P.A., Himelein, J.A., Paben, S.P., Ernst, L.R., and Eltinge, J. L.(2007), "Comparison of Simulation Methods Using Historical Data in the U.S. International Price Program", 2007 Proceedings of the American Statistical Association, Third International Conference on Establishment Surveys [CD-ROM], Alexandria, VA: American Statistical Association: to appear

[4] Efron, B. (1979). "Bootstrap methods: Another Look at the Jackknife." The Annals of Statistics. Vol 7, No. 1, pp. 1-26

[5] Efron, B. (2000). "The Bootstrap and Modern Statistics", Journal of American Statistical Association, Vol 95, No. 452, pp. 1293-1296

[6] Efron, B. and Gong, G. (1983). "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation." The American Statistician. Vol. 37, No. 1, 36-48.

[7] Efron, B., and Tibshirani, R. (1986). "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy", Statistical Science, Vol 1, No. 1, pp. 54-75

[8] Efron, B. and Tibshirani, R.J. (1993). "An Introduction to the Bootstrap." Chapman & Hall.

[9] Korn, E. and Graubard, B. (1990). "Simultaneous testing of regression coefficients with complex survey data: Use of Bonferroni t statistics." The American Statistician, 44, 270-276.

[10] Lahiri, P. (2003). "On the impact of bootstrap in survey sampling and small-area estimation." Statistical Science, Vol.18, No. 2, 199-210.

[11] Rao, J.N.K. and Wu, C.F.J. (1988). "Resampling inference with complex survey." Journal of American Statistical Association, Vol.83, No. 401, 231-241.

[12] Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). "Some recent work on Resampling methods for complex surveys." Survey Methodology, Vol.18, No. 2, 209-217.

[13] Sitter, R. (1992). "A resampling Procedure for Complex Survey Data." Journal of American Statistical Association, Vol 87, No. 419, pp. 755-765

[14] Toftness, Richard, "Comparison of Variance Estimation Techniques for a Price Index with Non-Independent Weights" 2002 Proceedings of the American Statistical Association. *Government Statistics Section [CD-ROM]*, (2002).

[15] Wolter, K.M., *Introduction to Variance Estimation.* Springer-Verlag, New York 1985.