Assessing Nonresponse Bias in the International Price Program's (IPP) Import and Export Price Index Surveys

Jenny FitzGerald

Bureau of Labor Statistics

2 Massachusetts Avenue NE, Washington DC 20212 U.S.A.

fitzgerald.jenny@bls.gov

Abstract

The IPP's Import and Export Price Index Surveys provide estimates of the change in prices of goods and services purchased by U.S. residents from foreign merchants or sold to foreign buyers by U.S. residents. In order to meet new Office of Management and Budget (OMB) standards and guidelines for statistical surveys, research was conducted using a three-phase approach to assess the effect of nonresponse bias in the IPP price indexes. The first phase modeled response rates using logistic regression. The second phase used a generalized linear mixed model with a random intercept term to investigate item price trends with varied response rates and company and item characteristics. Finally, using an equation provided by OMB to estimate nonresponse bias, we determined whether the level of nonresponse error in the price indexes was significant.

KEY WORDS: Nonresponse Bias; Logistic Regression; Generalized Linear Mixed Model with Random Intercept; Nonresponse Adjusted Sampling Weights; Wilcoxon Signed-Rank Test

1. Introduction

The International Price Program (IPP) produces import and export price indexes that measure the change in prices of goods and services purchased by U.S. residents from foreign merchants or sold to foreign buyers by U.S. residents. In 2006 the Office of Management and Budget (OMB) released a set of "Standards and Guidelines for Statistical Surveys." Within the standards, OMB mandated that any federal survey with a unit response rate below 80 percent (Guideline 3.2.9) or item response rate below 70 percent (Guideline 3.2.10) conduct a nonresponse analysis.

From 2003 to 2008, the establishment response rate for the IPP Import Price Index Survey varied between 84 percent and 89 percent; while the establishment response rate for the IPP Export Price Index Survey varied between 82 percent and 86 percent. Even though the establishment response rates met OMB's minimum requirement of 80 percent, the same cannot be said about the response rates at the item level. That is, the item response rate for the import survey ranged from 65 percent to 71 percent from 2003 to 2008, and the item response rate for the export survey ranged from 59 percent to 68 percent. Consequently, the IPP conducted research using a three-phase approach to assess the effect of nonresponse bias on its import and export price indexes.

This paper discusses the three individual phases of the IPP's Nonresponse Bias Study. First, a brief overview of the IPP sample design and index estimation process is provided. Next, the study's methodology is described. Finally, the results from each phase are presented.

2. Sampling in the IPP

The IPP's import merchandise sampling frames are obtained from the U.S. Customs and Border Protection (USCBP). The export merchandise sampling frames are from the Canadian Customs Service for exports to Canada and from the U.S. Bureau of the Census. The frames contain information about all import and export 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.

The IPP divides the import and export universes into two halves referred to as panels (Bureau of Labor Statistics, 2007). One import panel and one export panel are sampled each year and sent to the field offices for collection, so that both universes are 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

The 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 ratio (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 (Bobbitt, 2007).

$$LTR_t = \left(\frac{\sum p_{i,t} q_{i,0}}{\sum p_{i,0} q_{i,0}} \right)(100)$$

$$= \left(\frac{\sum p_{i,0} q_{i,0} \left(\frac{p_{i,t}}{p_{i,0}} \right)}{\sum p_{i,0} q_{i,0}} \right)(100)$$

$$= \left(\frac{\sum w_{i,0} r_{i,t}}{\sum w_{i,0}} \right)(100)$$

$$= \left(\frac{\sum w_{i,0} r_{i,t}}{\sum w_{i,0} r_{i,t-1}} \right)\left(\frac{\sum w_{i,0} r_{i,t-1}}{\sum w_{i,0}} \right)(100)$$

$$= \left(\frac{\sum w_{i,0} r_{i,t}}{\sum w_{i,0} r_{i,t-1}} \right)(LTR_{t-1})$$

$$= (STR_t)(LTR_{t-1})$$

where,

$p_{i,t} = $ price of item i at time t

$q_{i,o} = $ quantity of item i in base period 0

$w_{i,0} = p_{i,0} q_{i,0}$, the total revenue in base period 0

$r_{i,t} = \dfrac{p_{i,t}}{p_{i,0}}$, or the long term ratio of item i at time t

$LTR_t = $ long term ratio of a collection of items at time t

$$STR_t = \left(\frac{\sum w_{i,0} r_{i,t}}{\sum w_{i,0} r_{i,t-1}} \right)$$

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 uses items from as many as three samples 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. Multiple items can be selected from the same company-classification group pair in more

than one sample. Thus, 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. Nonresponse Bias Project

The IPP chartered the Nonresponse Bias Project to assess the effect of nonresponse bias on its import and export price indexes. The IPP collects price data from sampled establishments at initiation and during monthly re-pricing. Accordingly, there are two possible sources of nonresponse error: initiation nonresponse and re-pricing nonresponse. To compensate for re-pricing nonresponse, the IPP imputes the missing prices of items each month. In order to increase initiation response rates, the IPP periodically reduces the number of items it asks the respondent to reprice or decreases the frequency of repricing requests. For example, the IPP might allow a reluctant respondent to reprice quarterly instead of monthly. The program, however, does not impute prices or make adjustments to the estimation process for establishments that refuse to participate in the program's survey at initiation.

5. Methodology

The IPP used a three-phase approach to analyze nonresponse error incurred at initiation. The first phase of the study identified which variables significantly affect the likelihood of a response. To determine whether or not response rates are random at the establishment-quote level, the IPP looked at the relationship between unweighted response rates and five auxiliary frame variables. Response rates at the establishment-quote level were calculated from data from four import samples and four export samples. Additionally, a logistic regression model was used to identify which variables significantly affect response rates. The model was run separately for imports and exports within each panel.

The second phase of the study determined if prices trend differently for company and item characteristics that demonstrated different response rates in the first phase. Two years of item short term ratios were examined. An item's short term ratio (STR) is the ratio of its current price compared to its price for the previous period.

The price change data was collected from four import samples and four export samples fielded between 2001 and 2004. Moreover, the data contained information from two import samples from each panel and two export samples from each panel. Variables that were found to be significant in the first two phases of the study were identified as the characteristics that put IPP indexes at risk of nonresponse bias.

The third phase of the study estimated the level of nonresponse error in the IPP's index estimates. To do this, adjusted sampling weights were calculated to compensate for initiation nonresponse for the variables found to be significant in the first two phases of the study. Then two sets of index estimates were calculated from the item relatives collected from January 2006 to January 2007 – first using the current sampling weights and a second time using the nonresponse adjusted sampling weights. Differences between the indexes were assumed to be biases.

5.1 Logistic Regression Model

Response is a binary outcome. As a result, the first phase of the study used a logistic regression model to identify which variables significantly affect response rates. The model was run separately for imports and exports within each panel. The logistic regression model is as follows:

$$\log\left[\frac{p_i}{1-p_i}\right] = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} \quad (1)$$

where:
$i = 1, …, $ n items;
$p_i = $ the probability of a response for item i ;
$k = $ the number of explanatory variables.

The logistic regression model assumes that the observations are independent and that the independent variables are linearly related to the logit as expressed in the above equation (Garson, 1998a).

Deviance and diagnostic plots of the deviance residuals against the predicted values were used to evaluate the fit of the regression models. Deviance is a measure of how well data are fitted by a statistical model. Deviance is computed as:

$$D = -2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{\hat{p}_i}{y_i}\right) + (1-y_i)\ln\left(\frac{1-\hat{p}_i}{1-y_i}\right)\right] \quad (2)$$

where,
$i = 1, …, $ n items,
$y_i = $ the observed response rate ,
$\hat{p}_i = $ the predicted response rate .

From Hosmer and Lemeshow (2000): the role that deviance plays for a logistic regression is the same role that the residual sum of squares plays in linear regression. Thus, a model where the deviance is less than or equal to the degrees of freedom is considered a good fit.

5.2 Generalized Linear Mixed Model

The second phase of the Nonresponse Bias Study modeled the relationship between item STRs and the predictor variables that were found to be significant in the first phase using Generalized Linear Mixed Models (GLMMs) with random item effects. Fixed effects models, which include linear models and logistic regression models, assume that all observations are independent of each other. Item STR data for the second phase of the study, however, were not independent of each other since the data consisted of repeated price measurements on items over time. Therefore, Sadler (2007a) selected the GLMM with random item effects since the modeling technique accounted for the correlation of the longitudinal data.

According to Charles E. McCulloch (2003), the idea behind GLMMs is to incorporate random item effects into the linear predictor portion of a generalized linear model. GLMMs extend the usual general linear model by the inclusion of random effects. A random intercept model adds a single random effect for each item to the linear predictor portion of a generalized linear model:

$$g(\mu_{it}) = \text{logit}(\mu_{it}) = \log\left[\frac{\mu_{it}}{1 - \mu_{it}}\right] = x'_{it}\beta + \alpha_i \quad (3)$$

where:

$i = 1,...,n$ items;

$t = 1,...,n_i$ repeated observations of STRs nested within each item;

$\mu_{it} = P(Y_{it} = 1 \mid \alpha_i, x_{it})$ the conditional probability of a response given the random effect for item i and the covariate values at time t where $Y_{it} = \begin{cases} 1 \text{ if a price change occurred} \\ 0 \text{ if no price change occurred} \end{cases}$;

$x_{it} = (x_{it1}, x_{it2},...,x_{itp})'$ is the $p \times 1$ vector of observed predictor variables of item i at time t ;

$\beta = (\beta_1, \beta_2,...,\beta_p)'$ is the set of p fixed effects;

$\alpha_i =$ random item effect.

The fixed effects were the frame variables. The random effects were item specific and represented the influence that each item had over its price change that was not captured by the frame variables. According to Everitt and Howell (2005), the random effects are treated as such since the sampled items are thought to represent a population of items that are normally distributed with mean 0 and variance σ_v^2.

Three different sets of GLMMs were investigated using the GLIMMIX procedure in SAS (SAS Institute Inc., 2008). The first model looked to see which frame variables significantly affect price change. The response variable was coded as one if a price change occurred and zero if no price change occurred. The second model examined which frame variables significantly affect the direction of a price change. Observations with no price change were removed. The response variable was coded as one for a price increase, and zero for a price decrease. The third model explored which frame variables are significant in explaining the magnitude of a price change. Once again, observations with no price change were removed, and magnitude of a price change was modeled as a lognormal variable:

$$g(\mu_{it}) = \mu_{it} = x'_{it}\beta + \alpha_i \quad (4)$$

where:

$i = 1,...,n$ items;

$t = 1,...,n_i$ repeated observations of STRs nested within each item;

$\mu_{it} = \log|STR_{it} - 1|$;

$x_{it} = (x_{it1}, x_{it2},...,x_{itp})'$ is the $p \times 1$ vector of observed predictor variables of item i at time t ;

$\beta = (\beta_1, \beta_2,...,\beta_p)'$ is the set of p fixed effects;

$\alpha_i =$ random item effect.

Each covariate from the regression output was evaluated to determine which ones were significantly correlated with the dependent variables. A stepwise procedure was then used to isolate which variables were important to the model using the partial F-test. According to David Garson (1998b), individual effects of each independent can be assessed by running a model with and without a given independent, then taking the partial F to test the difference.

5.3 OMB Nonresponse Bias Equation

OMB provided the following equation to estimate nonresponse bias:

$$B(\bar{y}_r) = \bar{y}_r - \bar{y}_t = \left(\frac{n_{nr}}{n}\right)(\bar{y}_r - \bar{y}_{nr}) \quad (5)$$

where:

$\bar{y}_t$ = mean based on all sample cases;

$\bar{y}_r$ = mean based only on respondent cases;

$\bar{y}_{nr}$ = mean based only on nonrespondent cases;

n = number of cases in the sample;

n_{nr} = number of nonrespondent cases.

By the definition of nonresponse, the price relatives of items from nonresponding establishments were not available. As a result, adjusted sampling weights were calculated based on the variables found to be significant in Phases I and II. The STRs calculated from the adjusted sampling weights were assumed to be approximately equal to the STRs of all sampled establishment-quote pairs. Therefore, the level of nonresponse bias was estimated as the difference between the mean calculated using the current sampling weights and the mean calculated using the adjusted sampling weights.

5.4 Significance Test

Wilcoxon signed-rank tests were conducted to see if the differences between the STRs calculated using the two sets of weights were significant. The Wilcoxon signed-rank test is a nonparametric alternative to the paired student's t-test. To be precise, data must be normally distributed for a paired t-test but not for the Wilcoxon signed-rank test. From Hollander and Wolfe (1999), the signed-rank test is designed for analyses in which the primary interest is centered on the median of a population. It was hypothesized that there would be no difference between the indexes calculated using the current sampling weights and the indexes calculated using the nonresponse adjusted weights. Thus, the median of the differences was expected to be zero.

6. Findings

6.1 Overview

Variables with a significant effect on response rates and item price trends are the characteristics that put the IPP indexes at risk of nonresponse error. Table 1 defines the

five auxiliary frame variables of interest. The next three sections discuss the findings from each phase of the Nonresponse Bias Study.

Variable	Description
Consistency Rank	A measure of the frequency of trade of a sampling unit based on the number of months and quarters an establishment traded a good during the reference period. The scale for consistency rank ranges from one to seven.
Number of Quotes Requested	The number of quotes requested from the respondent when the sample was fielded. The number of quotes requested was divided into five classes.
Dollar Weight	The dollar value of trade represented by a sampling unit. Dollar weight was transformed from a quantitative variable into a categorical variable with five levels, each level representing a particular quintile.
Collection Region	The regional office that had jurisdiction over the state in which the initiation was performed.
Sampling Stratum	Product area represented by the two-digit section area from the Harmonized Classification system.

Table 1: Auxiliary Frame Variables

## 6.2 Results from Logistic Regression Model

From Sadler (2006), the logistic regression model from the first phase indicated that consistency rank, number of quotes requested, dollar weight, region, and sampling strata all significantly affect response rates at the $\alpha = 0.05$ level.

Consistency rank significantly affected response rates across panels for import and export samples ($p \leq 0.001$). As consistency rank increased, nonresponse rates decreased. Graph 1 depicts the negative correlation that exists between nonresponse rate and consistency rank for Panel A imports.
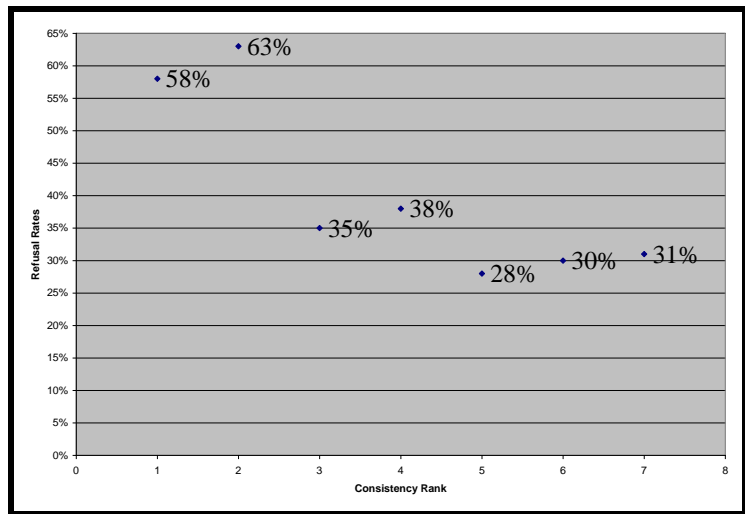
Graph 1: Nonresponse Rate by Consistency Rank

Number of quotes requested had the most significant effect on response rates in comparison to the four other frame variables. The number of quotes requested was significant at all five levels for the Panel A export sample ($p < 0.0001$) and the Panel B import and export samples ($p \leq 0.0144$). For the Panel A import sample, the number of quotes requested was significant at levels 1, 2, and 3 ($p \leq 0.0092$). As the number of quotes requested increased, nonresponse rates increased accordingly. This effect was especially pronounced when the number of quotes requested exceeded 15. Graph 2 shows the nonresponse rate for Panel A imports by the number of quotes requested.
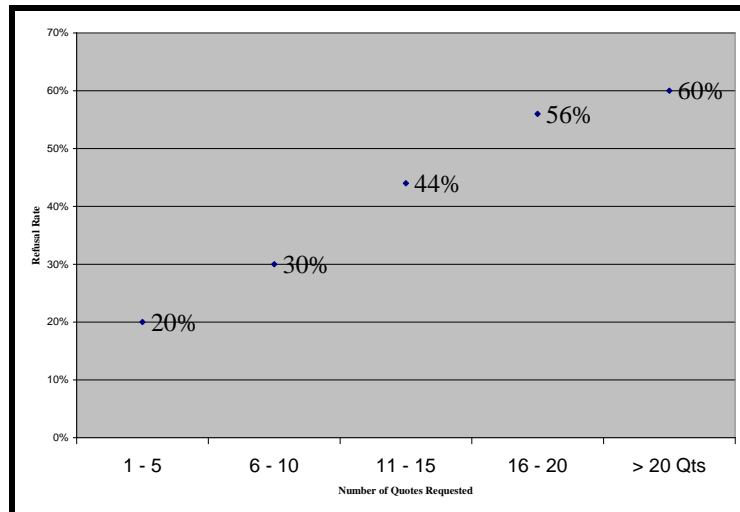
Graph 2: Nonresponse Rate by the Number of Quotes Requested

Dollar weight, collection region, and sampling stratum also had significant affects on response rates ($p \leq 0.0296$).

Diagnostic plots of the deviance residuals versus the predicted values showed that the model fit all four sets of data well. The deviance values, on the other hand, indicated that the fits could be improved. The ratio of the deviance value compared to the degrees of freedom ranged from 1.1668 to 1.2697 for the four sets of data (Sadler, 2006). Thus, the deviance was greater than the degrees of freedom for all four regressions. Nonetheless, the primary objective of the first phase of the study was achieved: to assess the relationship between response rates and the five auxiliary frame variables.

6.3 Results from GLMMs

The second phase of the study examined whether company and item characteristics with different response rates have an effect on item price changes. The five variables of interest were the ones found to be significant in modeling response rates in Phase I.

Three different sets of Generalized Linear Mixed Models with fixed item effects were investigated. The first model looked to see which frame variables significantly affect price changes. The second model examined which frame variables significantly affect the direction of a price change. Finally, the third model investigated which frame variables are significant in explaining the magnitude of a price change.

From Sadler (2007a): dollar weight, number of quotes requested, and sampling stratum were all significant in explaining variations in price change in the first model ($p < 0.05$). Additionally, dollar weight, number of quotes requested, and sampling stratum were all important to the models from the partial F-Tests ($p < 0.05$). Consistency rank was the only variable that was not important in explaining price change.

Model 2 found that number of quotes requested, dollar weight, and sampling stratum were significant to the variations in price increases and were important in the model ($p < 0.05$). Consistency rank, dollar weight, and region had some values that were significant to variations in price increases, but were not important to the model from the partial F-Tests ($p > 0.05$).

Model 3 revealed that only sampling stratum was significant in explaining variations in the magnitude of a price change ($p < 0.05$). Sampling stratum was also important in the model after all of the variables were added.

The ratio of the generalized chi-square statistic to the degrees of freedom was evaluated to see how well each model fit the data. Ratio values close to one generally indicate a good fit. For Model 1, the ratio of the generalized chi-square statistic to the degrees of freedom ranged from 0.61 to 0.68. For Model 2 the ratio ranged from 0.98 to just slightly less than one. For Model 3, the ratio ranged from 2.15 to 5.75, suggesting over dispersion. Overall the ratios of the generalized chi-square statistic to the degrees of freedom and diagnostic plots of residuals versus predicted values showed that all three responses were modeled appropriately.

6.4 Estimate of Nonresponse Bias

The third phase of the Nonresponse Bias Study estimated the level of nonresponse error in the IPP's import and export price indexes. By the definition of nonresponse, not all price relatives from all sampled establishment-quote pairs were available. As a result, nonresponse adjustment factors were calculated for the sampling weights for Stages I and II based on sampling stratum. Sampling stratum was the only variable from Phase II to affect the occurrence, direction, and magnitude of a price change.

6.4.1 Adjustment Factors for Stage I Weights

An establishment's Stage I weight is equal to its dollar value divided by its probability of being selected within a sampling stratum. Sadler (2007b) calculated the nonresponse adjustment factors for Stage I from the Stage I weights of responding and nonresponding establishments. To be precise, the nonresponse adjustment factors for Stage I were set equal to the Stage I weight of responding and nonresponding establishments within a sampling stratum divided by the Stage I weight of responding establishments only. Consequently, the adjustment factor inflated the old Stage I weight by the proportion of missing weight.

$$NRF1 = \frac{\sum_{i \in R} \text{Stage\_1\_Wg t}_i + \sum_{i \in N} \text{Stage\_1\_Wg t}_i}{\sum_{i \in R} \text{Stage\_1\_Wg t}_i} \quad (6)$$

where,
i is an establishment;
R is the set of respondents within a sampling stratum; and
N is the set of nonrespondents within a sampling stratum.

6.4.2 Adjustment Factors for Stage II Weights

A second set of nonresponse adjustment cells were created based on responding establishments within each sampling stratum. The weight for an establishment-classification group pair for Stage II is equal to its dollar value divided by the probability of the classification group being selected in the second stage of sampling.

The nonresponse adjustment factors for Stage II (NRF2) were set equal to the Stage II weight of responding and nonresponding quotes within an establishment-stratum divided by the Stage II weight of responding quotes within the establishment-stratum (Sadler, 2007b). Consequently, the adjustment factor inflated the old Stage II weight by the proportion of missing weight.

$$NRF2 = \frac{\sum_{ij \in R} \text{Stage\_2\_Wgt}_i + \sum_{ij \in N} \text{Stage\_2\_Wgt}_i}{\sum_{ij \in R} \text{Stage\_2\_Wgt}_i} \quad (7)$$

where,
i is an Establishment;
j is a quote;
R is the set of respondent quotes within a responding establishment-stratum; and
N is the set of nonrespondent quotes within a responding establishment-stratum.

A nonresponse adjustment factor of one for Stages I and II occurred when nonresponse was zero, and an adjustment factor of two occurred when the magnitudes of response and nonresponse were equal. No constraints were placed on the adjustment factors. That is to say, survey managers often place an upper limit on weighting adjustments so that one response does not dominate a survey estimate. In order to observe the most extreme cases, no limits were placed on the nonresponse adjustment factors.

6.4.3 Results from Wilcoxon Signed-Rank Tests

The Wilcoxon signed-rank tests indicated that there was no significant difference between the two sets of indexes for 93 percent of the published import indexes and 96 percent of the published export indexes at the $\alpha = 0.05$ level (FitzGerald and Slusher, 2008). The indexes that might include nonresponse bias all differed by less than one percent. In regard to all of the indexes, about 61 percent differed by 0.1 percent or less; 94 percent differed by 0.2 percent or less; and 97 percent differed by 0.4 percent or less. Since the IPP only publishes its indexes to a tenth of a percent, these differences are negligible.

7. Conclusion

The nonresponse adjusted sampling weights did not significantly impact the IPP index estimates. The minimal impact from the adjusted sampling weights is thought to be the

result of the fixed aggregation weights at the classification group level. Furthermore, nonresponse bias does not appear to be an issue for the IPP index estimates.

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

References

Bobbitt, P., Paben, S., Cho, M.J., Chen, T., Himelein, J. and L. Ernst, (2007). "Application of the Bootstrap Method in the International Price Program." *Proceedings on the Section on Survey Methods Research, American Statistical Association* [CD ROM], 2910-2917.

Bureau of Labor Statistics (2007). "Chapter 15 of the BLS Handbook of Methods: International Price Indexes." Accessed Oct. 31st, 2008 from the Bureau of Labor Statistics Website http://www.bls.gov/opub/hom/homtoc.htm.

Everitt, Brian and David C. Howell (eds.) "Generalized Linear Mixed Models," *Encyclopedia of Statistics in Behavioral Science*. Vol. 4. John Wiley & Sons, Ltd., Chichester 2005.

FitzGerald, Jenny and Daryl Slusher (2008). "The Impact of Sample Adjusted Weights on the IPP Import and Export Price Indexes: Part IV of the Nonresponse Bias Study." Internal BLS Report dated September 30th, 2008.

Garson, G. David (1998a). "Logistic Regression." *Quantitative Research in Public Administration*. North Carolina State University's (NCSU). Accessed December 2006. http://faculty.chass.ncsu.edu/garson/PA765/logistic.htm#biblio.

Garson, G. David (1998b). "Multiple Regression." *Quantitative Research in Public Administration*. North Carolina State University's (NCSU). Accessed March 2009. *http://faculty.chass.ncsu.edu/garson/PA765/regress.htm*.

Hollander, Myles and Douglas A. Wolfe, (1999). Nonparametric Statistical Methods, Second Edition, New York: John Wiley, pp. 35-42.

Hosmer, David W. and Stanley Lemeshow. Applied Logistic Regression, Second Edition. John Wiley & Sons, New York (2000).

McCulloch, Charles E. "Generalized Linear Mixed Models" *NSF-CBMS Regional Conference Series in Probability Statistics*, Vol. 7, (2003).

Office of Management and Budget (2001). "Measuring and Reporting Sources of Error in Surveys." The Federal Committee on Statistical Methodology. Accessed July 15, 2008. http://www.fcsm.gov/01papers/SPWP31_final.pdf.

Office of Management and Budget (2006). "Standards and Guidelines for Statistical Surveys." Accessed Sep. 16, 2008. http://www.whitehouse.gov/omb/inforeg/statpolicy/standards_stat_surveys.pdf.

Sadler, Andy (2006). "Nonresponse Study Part I: Modeling Establishment/Quote Refusal Rates Using Logistic Regression." Internal BLS Memorandum dated December 6th, 2006.

Sadler, Andy (2007a). "Nonresponse Study Part II: Modeling Short Term Ratios Using a Generalized Mixed Model with Random Intercept Term." Internal BLS Memorandum dated June 6th, 2007.

Sadler, Andy (2007b). "Nonresponse Study Part III: Calculating a Set of Nonresponse Adjusted Sampling Weights." Internal BLS Memorandum dated December 13, 2007.

SAS Institute Inc. (2008). SAS/STAT 9.1 User's Guide: The Glimmix Procedure. Cary, NC: SAS Institute Inc.