Analyzing Nonresponse Bias in PPI Data

Boriana Chopova, Teresa E. Hesley, Greg Kelly, Soon Paik, Andy Sadler, Dave Slack
U.S. Bureau of Labor Statistics
2 Massachusetts Avenue NE, Washington DC 20212 U.S.A.
sadler.andy@bls.gov

Abstract
The Producer Price Index (PPI) program conducted a study to determine if nonresponse bias exists in PPI data. The study investigated nonresponse at unit initiation (when units are asked to participate in PPI samples) and during the unit/item repricing period (when units provide prices for the items they've agreed to reprice at initiation). The study consisted of three stages: In the first stage a contingency table analysis and regression models were used to analyze the relationship between response and several frame variables. In the second stage a regression model was used to analyze the relationship between item short term relatives and the variables which were associated with response from the first stage. The third stage tested the impact of sample adjusted weights on PPI indexes. This paper reports the results of this study.

Key Words: Nonresponse bias, unit initiation, item repricing, item short term relative, sample adjusted weights

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

1. Introduction

The Producer Price Index (PPI) program undertook a study to analyze the effects of nonresponse on its price index estimates. The study was initiated to comply with the Office of Management and Budget's mandate that federal surveys whose response rates fall below certain thresholds (80% for establishments and 70% for items) should conduct a nonresponse bias analysis. PPI establishment initiation survey response rates consistently range between 83 and 84 percent while those for item repricing response consistently settle around 66 percent. As is evident, the PPI response rates indicate a nonresponse bias study would be useful.

The Producer Price Index (PPI) of the Bureau of Labor Statistics (BLS) is a family of indexes that measure the average change over time in the prices received by domestic producers of goods and services. PPIs measure price change from the perspective of the seller. More than 100,000 price quotations per month are organized into three sets of PPIs: (1) Stage-of-processing indexes, (2) commodity indexes, and (3) indexes for the net output of industries and their products. The stage-of processing structure organizes products by class of buyer and degree of fabrication. The commodity structure organizes products by similarity of end use or material composition. The entire output of various industries is sampled to derive price indexes for the net output of industries and their products. PPIs for the net output of industries and their products are grouped according to the North American Industry Classification System (NAICS).

2. Sampling

The PPI typically uses the BLS sample and research database known as the Longitudinal Database (LDB) as the source of frame information for most of the industries sampled. The LDB contains U.S. business frame records representing all U.S. non-farm industries, with the exception of some sole proprietors. The LDB consists of all covered employers under the Unemployment Insurance (UI) Tax System. The frame information used to cluster establishments on the LDB is the Employer Identification Number (EIN).

The 6-digit NAICS industries are sampled using a two-stage design. First-stage sample units are selected in the Washington office from a list of establishments and clusters of establishments whose primary production is thought to be in a given 6-digit NAICS industry. The final or second-stage sample units are then selected during data collection at the location of the sampled establishment. The second-stage units are unique items, products, or services for which the respondent is to report prices monthly for 5-7 years.
The first-stage sample units are selected systematically with probability proportional to a measure of size. The measure of size is usually employment when the Longitudinal Database is used as the frame source for sampling. The measure of size is thought to correlate with revenue, which is collected directly from a sampled unit and used in weights of items in index calculation. The second-stage sample units are selected in the field at the location of the establishment selected in the first stage.

3. PPI Response Rates

Several response rates are calculated in the PPI but we focused our research on the initiation response rate and the repricing response rate.

3.1 Initiation Response Rate

The initiation response rate measures the proportion of sampled units, sent to the field for collection, which agree to participate in PPI surveys. The initiation response rate is the ratio of the number of productive units (i.e., those with a productive status code) to the number of productive units, refusal units (those with a refusal status code) and unknowns (those with unknown status codes). The calculation is done as follows:

$$\text{Initiation Response Rate (Unweighted)} = \frac{\sum_{All\ Units} Productive\ Units}{\sum_{All\ Units} Productive + Refusal + Unknown}.$$

3.2 Repricing Response Rate

The repricing response rate measures the proportion of items requested from respondents which are being used in index estimation. The calculation is done as follows:

Repricing Response rate (unweighted) =

$$\frac{\sum_{All\ Units} All\ items\ in\ Estimation + Non\ index\ items\ with\ good\ prices + Mailed\ Items}{\sum_{All\ Units} All\ mailed\ items + Non\ mailed\ items\ with\ reported\ data}.$$

4. Modeling Initiation and Repricing Response

Data used in the analysis of initiation nonresponse consisted of both responding and nonresponding units. Data used in repricing nonresponse used only responding units. The explanatory variables used in our initiation response models were Collected Region and Employment. For the repricing models, Shipments and Receipts was used instead of Employment. We used a logistic regression model to analyze unit initiation response and a survival analysis model to analyze unit and item repricing nonresponse.

4.1 Initiation Nonresponse

Initiation response measures unit response. Our dependent variable in the model was unit initiation response. There were two possible response states: cooperative or refusal response. A unit was classified as cooperative at initiation if it provided at least one item for repricing. Otherwise, the unit was classified as a refusal. We identified all available explanatory variables and analyzed them. We modeled correlation between the available variables and chose the most appropriate ones among those that showed high correlation with each other.

4.1.1 Data

The data consisted of industries where employment was used as the sampling measure of size. We used industries that were sampled as NAICS and had two years of repricing history which limited our study to industries introduced between 2004 and 2008. The explanatory variables used were Collected Region and Shipments and Receipts.

Region refers to the region where the unit was collected. There are six regions not including the National Office. Units which were collected by the National Office were re-classified and placed in the region of their headquarters location.

Table 1. Response rates by region

		Response			
Region	Number of Units	Cooperative Units	Refusal Units	Percent Cooperative	Percent Refusal
Mid West	3752	3306	446	88%	12%
West	2655	2271	384	86%	14%
South East	2828	2429	399	86%	14%
North East	2141	1841	300	86%	14%
Mid Atlantic	2290	1920	370	84%	16%
South	2371	2003	368	84%	16%

Employment was used as a measure of unit size. To aid in the analysis, we coded employment (across all industries) as a categorical variable with 5 levels corresponding to quintiles of employment.

Table 2. Response rates by employment

		Response			
Employment Quintiles	Number of units	Cooperative Units	Refusal Units	Percent Cooperative	Percent Refusal
(1 <= Emp <= 22)	3257	2740	517	84%	16%
(22 < Emp <= 61)	3175	2747	428	87%	13%
(61 < Emp <= 146)	3215	2794	421	87%	13%
(146 < Emp <= 394)	3190	2767	423	87%	13%
(394 < Emp)	3200	2722	478	85%	15%

4.1.2 Methodology

This methodology employed for modeling unit initiation response is a logistic regression model. In a logistic regression model the outcome variable (y) is binary (0 or 1). Logit models predict the probability (p) that y = 1. Our generic logistic model in linear form for k explanatory variables and $i = 1,\ldots,n$ observations looks for each variable like this:

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

where p_i is the probability that response $y_i = 1$. The expression

$$\frac{p}{1-p} = \frac{Probability\ of\ event}{Probability\ of\ no\ event}$$

is called the odds of an event. The odds have a lower bound of 0 but no upper bound. The log of the odds is called the *logit* or *log-odds*. We may solve for p_i in the logit equation to get:

$$p_i = \frac{exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})}{1 + exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})} \quad (2)$$

$exp(x)$ [also written e^x] is the exponential function where $e \cong 2.71824$ and $log(e^x) = x$. From equation (1) we may interpret the results of our regressions by calculating the odds of a positive response and from equation (2) we may interpret the regression results by calculating the probability of a positive response. Interpretation of logistic regression results in terms of odds is generally more common. To calculate the odds ratios, we exponentiate the estimated coefficients of the logit model. For a logistic regression model with a dichotomous independent variable coded 0 and 1, the relationship between the odds ratio and the estimated regression coefficient is: $\widehat{Odds\ Ratio} = exp(\hat{\beta})$.
For a categorical explanatory variable with two levels, for example, the odds ratio is a measure of association which approximates how much more likely or unlikely for the outcome to occur for those with x = 1 than for those with x = 0. (x = 0 is our baseline or reference group to which all other categories/levels of x will be compared in terms of their propensity to respond).
The logistic regression model assumes that the observations are independent and that the independent variables are linearly related to the logit as expressed in equation (1). Notice we are modeling the probability that unit response $y = 1$, i.e., that the unit has agreed to reprice at least one item requested at initiation.

To measure how well data are fitted by the statistical logit model we calculate the deviance.[1]

$$D = -2\sum_{i=1}^{n}\left[y_i ln\left(\frac{\hat{p}_i}{y_i}\right) + (1-y_i) ln\left(\frac{1-\hat{p}_i}{1-y_i}\right)\right] \quad (3)$$

where $\hat{p}_i = \hat{p}_{(x_i)}$. According to Hosmer and Lemeshow, the deviance for logistic regression plays the same role that the residual sum of squares plays in linear regression. Values of the deviance equal to or less than the degrees of freedom usually indicate a good fit.

4.1.3 Results

Table 3 below shows the results for the deviance calculated for assessing goodness of fit. The value of the deviance for this model is less than the degrees of freedom, which indicates, according to Hosmer and Lemeshow, a good fit.

[1] See Hosmer and Lemeshow, *Applied Logistic Regression*, p.13.

Table 3. Criteria for assessing goodness of fit

Criterion	DF	Value	Value/DF
Deviance	16 E3	13025.4369	0.8127

Table 4 shows the results of global significance tests for the null hypothesis that all the coefficients for an explanatory variable are equal to zero. In Table 4 we see that both Region and Employment are significantly correlated with unit initiation response.

Table 4. Likelihood ratio statistics

Source	DF	Chi-Square	Pr > ChiSq
Region	5	26.70	<.0001
Employment	4	14.61	0.0056

In Table 5, we display results for the estimated values of the parameters in our model along with 95% confidence intervals and estimated odds ratios.

Table 5. Analysis of parameter estimates for unit initiation response

Parameter	DF	Estimate	SE	95% Confidence interval		Chi-Square	Pr > ChiSq	Estimated odds ratio $[e^{\hat{\beta}}]$
Intercept	1	1.7050	0.0721	1.5638	1.8463	559.61	<.0001	
REGION								
North East	1	0.0368	0.0832	-0.1264	0.1999	0.20	0.6588	1.037486
Mid Atlantic	1	-0.1289	0.0793	-0.2843	0.0265	2.64	0.1040	0.879062
South East	1	0.0289	0.0774	-0.1227	0.1805	0.14	0.7086	1.029322
Mid West	**1**	**0.2218**	**0.0750**	**0.0747**	**0.3688**	**8.74**	**0.0031**	**1.248322**
South	1	-0.0823	0.0792	-0.2376	0.0729	1.08	0.2984	0.920996
West	0	0.0000	0.0000	0.0000	0.0000	.	.	.
EMPLOYMENT								
Emp_1	1	-0.0522	0.0693	-0.1881	0.0836	0.57	0.4511	0.949139
Emp_2	1	0.1301	0.0721	-0.0112	0.2714	3.26	0.0711	1.138942
Emp_3	**1**	**0.1618**	**0.0722**	**0.0203**	**0.3033**	**5.02**	**0.0251**	**1.175625**
Emp_4	1	0.1407	0.0721	-0.0006	0.2820	3.81	0.510	1.151079
Emp_5	0	0.0000	0.0000	0.0000	0.0000	.	.	.

In the table we see that:

- The only level of Region which is significant is the Midwest. The odds of a positive response by units in the Mid West region are 24.8% higher than for units in the West region (the baseline region).
- Employment is significant only at level 3. The odds of a positive response by units at level 3 of employment are 17.5% greater than units at level 5 (the baseline employment level).

4.2 Unit and Item Repricing Nonresponse

Unit repricing response was modeled as the number of months a unit provided good prices for at least one of the items it agreed to reprice. Item repricing response was modeled as the number of months an item had a good price. The covariates used were

Region and Shipments and Receipts (S&R). S & R was thought to be a better measure of unit size than Employment.

The repricing response models used data from industries which entered estimation between 2004 and 2008 and were repriced for up to two years. The analysis modeled the number of months a unit/item provided good prices for index estimation. It used the first two years after index introduction as the analysis period. Units/items still providing good prices for index estimation beyond the two year period of the study were **censored**.[2]

4.2.1 Data

The following tables display the numbers of censored and uncensored units by Region and Shipments and Receipts across the industries and samples analyzed.

Table 6. Number of uncensored and censored units for Region

Region	Number of units	Number of uncensored units	Number censored units
Midwest	3692	1556	2136
Southeast	2608	1235	1373
West	2405	1248	1157
South	2182	1036	1146
Northeast	2092	1030	1062
Mid-atlantic	2053	927	1126

The West (52%) and NE (49%) had the highest rates of uncensored units while the Mid West (42%) and Mid Atlantic (45%) had the lowest rates.

Table 7. Number of uncensored and censored units for Shipments and Receipts

S & R quintile ($)	Number of units	Number of uncensored units	Number censored units
[250 <= S&R <=2,300,000]	3640	2086	1554
(2,300,000 < S&R <= 12,000,000)	3526	1592	1934
(12,000,000 < S&R <= 41,008,100)	3176	1356	1820
(41,008,100 < S&R <= 150,000,000)	2768	1140	1628
(150,000,000 < S&R)	1922	858	1064

The variable Shipments and Receipts is used as an indicator of unit size. The variable was categorized into five levels each representing 20% of the total data. The smallest units are included in the lowest quintiles (1, 2, 3) and the largest units in the highest quintiles (4 and 5). The highest percentage of uncensored units is seen in levels 1, 2 and 5 and the lowest percentage is seen in levels 3 and 4. Generally speaking, the table shows that smaller units stopped providing good prices at a higher rate than larger units.

[2] The term describes those units and items which provided more than two years of good prices. Units and items which provided two or more years of good prices were censored. Those that provided less than two years of good prices were uncensored.

4.2.2 Methodology

For the study of repricing nonresponse, we utilized both a Cox regression model as well as a Logistic regression model.

4.2.2.1 Cox Regression

Cox regression (survival analysis) methods apply whenever there is an interest in examining the time to occurrence of an event. The focus of these models is on the distribution of survival times. The unique feature of survival data is the presence of censored observations. In the present analysis, censoring occurred for units/Items which had provided good prices for use in index estimation more than two years after inception.

The relationship of survival times to the explanatory variables in the Cox model is written as:[3]

$$h_i(t) = h_o(t)\exp\,(\beta_i x_{it} + \ldots + \beta_i x_{ik})$$

Where:

$h_i(t)$ = the hazard function for the i^{th}unit/item
$h_o(t)$ = the baseline hazard
β = the vector of coefficients for the explanatory variables x_i in the model.

The results for the Cox regression model of unit survival in index estimation as a function of Shipments and Receipts and Region appear below.

Table 8. Testing global null hypothesis: BETA=0

Test	DF	Chi-Square	Pr > ChiSq
Likelihood Ratio	9	285.0903	<.0001
Score	9	299.1297	<.0001
Wald	9	296.1401	<.0001

Table 8 displays the results of the Likelihood Ratio, Score and Wald tests. The likelihood ratio chi-square test compares the log-likelihood for the fitted model to the log-likelihood for a model with no explanatory variables. All three results show that the model has significant explanatory power.

In Table 9, we display results for the estimated values of the parameters in our model along with estimated odds ratios.

[3] The hazard function may be interpreted as the rate at which units/items stop providing good prices for use in index estimation per month.

Table 9. Parameter estimates for unit repricing response

Variable	DF	Parameter Estimate	SE	Chi-Square	Pr > ChiSq	Hazard ratio
Shipments and Receipts						
1	1	0.34599	0.04058	72.6881	<.0001	1.413
2	1	0.01225	0.04238	0.0836	0.7725	1.012
3	1	-0.05946	0.04363	1.8570	0.1730	0.942
4	1	-0.10909	0.04523	5.8168	0.0159	0.897
5	0	0	.	.	.	.
Region						
NE	1	-0.09781	0.04210	5.3976	0.0202	0.907
Mid Atlantic	1	-0.19977	0.04339	21.1926	<.0001	0.819
SE	1	-0.13333	0.04017	11.0138	0.0009	0.875
Mid West	1	-0.28453	0.03806	55.8857	<.0001	0.752
South	1	-0.15219	0.04204	13.1058	0.0003	0.859
West	0	0	.	.	.	.

The results for the Cox regression of the number of months a unit provided good prices for use in index estimation show:

- Region is highly significant in explaining the number of months a unit provided good prices for use in index estimation. The West region has a higher hazard of items not providing good prices than all the other regions in the study.
- Shipments and Receipts is significant at two levels. For units in the lowest quintile (level 1) of Shipments and Receipts the hazard of a unit not providing good prices for at least one item to be used in index estimation is 41.3% higher than for units in the highest quintile (level 5). At level 4 of Shipments and Receipts, the hazard of a unit not providing good prices for the index is 89.7% of that at level 5 (I.e., level 4 has a lower hazard of not providing good prices than level 5).

4.2.2.2 Logistic Regression

In addition to modeling with Cox Regression, we also used Logistic Regression models to explore the significance of Region and Shipments and Receipts in unit and item repricing nonresponse. The results of these models are in agreement with the results of the Cox Regression models in that both explanatory variables are deemed to be significant.

5 Modeling Item Prices

We modeled the relationship between item short term relatives and the two variables which were found to be statistically significant in predicting response: Collected Region and unit Shipments and Receipts. We wanted to determine if price trends were strongly correlated to these variables. We modeled the behavior of item prices in three ways:

- Model I: As a binary variable indicating whether a price change occurred.
- Model II: As a binary variable indicating the direction of the price change.
- Model III: As a lognormal variable indicating the magnitude of the price change.

The term ‘item short term relative’ is used to describe a number which expresses a percent change in the price of an item from one time period to the next. Interest naturally centers on whether there's an association between price changes and certain characteristics of the sample in which those items were collected.

5.2 Data

The data used in our analysis were compiled for industries introduced between 2004 and 2008. Information on items (and their short term relatives) collected during the period from January, 2004 to January, 2010 was used. Our data contained 324 six digit NAICS industries of which 196 had sufficient numbers of observations to be used in the analyses. We created nine industry groups based on similar NAICS industry codes from the industries with sufficient data.[4]

5.3 Methodology

Our analysis used a Generalized Linear Mixed Model with random item effects, which account for the correlated structure of the data. The usual fixed effects models, or generalized linear models which include linear and logistic regression models, assume that all observations are independent of each other. Fixed effects models are in-appropriate for use with data which are correlated such as clustered data and longitudinal or repeated measures data. Our data consists of repeated measurements of short term relatives on several thousand items over time. The structure of our data (price changes measured for single items over a one year period) is highly correlated therefore justifying the use of a modeling technique which takes account of these correlations.

Let i denote the items and let j denote the short term relatives. Assume there are $i = 1, \cdots, N$ items and $j = 1, \ldots, n_i$ repeated observations nested within each item. A random-intercept model, the simplest mixed model, adds a single random effect for item i to the linear predictor: $\omega_{ij} = x'_{ij}\beta + \gamma_i$ where γ_i is the random effect (one for each item). These random effects represent the influence of item i on its repeated observations that is not captured by the observed covariates (the fixed effects). These are treated as random effects because the sampled items are thought to represent a population of items, and they are usually assumed to be distributed as $N(o, \sigma_\gamma^2)$[5]. The parameter σ_γ^2 indicates the variance in the population distribution, and therefore the degree of heterogeneity (differentness) of items. The random effects part of the model is a mechanism for representing how correlation occurs between observations within a cluster[6] of items.

The mixed-effects logistic regression model is a common choice for analysis of binary data. In the GLMM context, the model utilizes the logit link,

$$g(\mu_{ij}) = logit(\mu_{ij}) = log\left[\frac{\mu_{ij}}{1 - \mu_{ij}}\right] = x'_{ij}\beta + \gamma_i$$

Here, the conditional expectation $\mu_{ij} = E(Y_{ij}|\gamma_i, x_{ij})$ equals, $P(Y_{ij} = (1|\gamma_i, x_{ij})$ namely, the conditional probability of a response given the random effects and covariate values. For the log-normal distribution used in modeling the magnitude of change in item short term relatives, the link function used is the identity link, i.e., $g(\mu_{ij}) = x'_{ij}\beta + \gamma_i$.

---

[4] The industry groups were: Mining and Construction, Nondurable Manufacturing (31), Nondurable Manufacturing (32), Durable Manufacturing, Trade and Transportation, Finance and Infrastructure, Health, Amusement, and Repair.

[5] This is read as: The items are normally distributed with mean 0 and variance σ_γ^2.

[6] The term cluster is used to describe the set of prices for the same item.

5.4 Results

Model I is a mixed effects logistic regression model in which the dependent variable (item short term relatives in binary format) is regressed against Collected Region and unit Shipments and Receipts. The response was coded as 1 if a price change occurred and as 0 if no price change occurred. The model is a generalized linear mixed model with a random intercept term. The random intercepts (one for each item) are deviations from this fixed estimate. Estimates of the random intercepts were not calculated. Model II is also a mixed effects logistic regression model where the dependent variable (the direction of price change) is also regressed against Collected Region and unit Shipments and Receipts. In Model II the response is coded as '1' if there was a price rise and as '0' if the price declined. Model III is a mixed effects regression model where the dependent variable (the absolute value of the price change) is assumed to have a logistic distribution and the link function used is the identity link.

Table 10. Summary Model I findings

	Explanatory variable				
	Region		S & R		
Industry Group	$t-test^{*}$	$F-test^{**}$	$t-test^{*}$	$F-test^{**}$	$Model\ Fit^{***}$
Mining and Construction	✓	✓	✓	✓	.90
Nondurable Manufacturing (31)	✓	✓	✓	✓	.76
Nondurable Manufacturing (32)	✓	✓	✓	✓	.79
Durable Manufacturing	✓	✓	✓	✓	.77
Trade and Transportation	✓	✓	✓	✓	.71
Finance and Infrastructure	✓	✓	✓	✓	.64
Health	×	×	✓	✓	.73
Amusement	✓	✓	✓	×	.80
Repair	✓	✓	×	✓	.76

*t-tests significant at 5% or less for one or more levels of the variable.
** Partial F-test significant at 5% or less.
*** If $\frac{Gener,Chi-Square}{DF}$ is close to 1 then the model is considered a good fit.

The findings for the relationship between price changes and Region and Shipments and Receipts are these:

- At least one or more levels of the Region was significant in explaining price changes in eight (8) of the nine (9) industry groups -- see the t-test results. Region was also important in eight (8) of the nine (9) industry groupings as measured by the significance of F-tests. The only industry where Region was not significant or important in explaining price changes was the Health industry grouping.
- At least one or more levels of Shipments and Receipts were significant in explaining price changes in all eight (8) of the nine (9) industry groups. Shipments and Receipts was also important in eight (8) of the nine (9) industry groups as measured by the F-test. Shipments and Receipts was not important in

explaining price changes in the Amusement industry group and no levels of this variable were significant in the Repair industry group.

- The general tendency observed is for prices changes to occur with greater odds and higher probability for larger units than for smaller units.

Table 11. Summary Model II findings

Industry	Explanatory variable				
	Region		S & R		
	$t-test^{*}$	$F-test^{**}$	$t-test^{*}$	$F-test^{**}$	$Model\ Fit^{***}$
Mining and Construction	✓	✓	✓	✓	.88
Nondurable Manufacturing (31)	✓	✓	✓	✓	.83
Nondurable Manufacturing (32)	✓	✓	✓	✓	.81
Durable Manufacturing	✓	✓	✓	×	.77
Trade and Transportation	✓	✓	✓	✓	.80
Finance and Infrastructure	✓	✓	✓	✓	.68
Health	✓	✓	✓	✓	.72
Amusement	✓	✓	✓	✓	.63
Repair	✓	✓	✓	✓	.83

*t-tests significant at 5% or less for one or more levels of the variable.
** Partial F-test significant at 5% or less.
*** If $\frac{Gener, Chi-Square}{DF}$ is close to 1 then the model is considered a good fit.

The findings for the relationship between price rises and Region and Shipments and Receipts are these:

- All of the Regions were significant in explaining price increases in all nine (9) industry groups. Region was also important in explaining price increases in all nine (9) industry groups as measured by the significance of F-tests.
- All of the levels of Shipments and Receipts were significant in explaining price increases in all nine (9) industry groups. Shipments and Receipts was also important in eight (8) of the nine (9) industry groups as measured by the F-test. Shipments and Receipts was not important in explaining price increases in the Nondurable Manufacturing (industry group 33).
- The general tendency observed is for prices rises to occur with lower odds and lower probability for smaller units than for larger units. The trend in Mining and Construction, and Healthcare was the reverse.

Table 12. Summary Model III findings

Industry	Explanatory variable				
	Region		S & R		
	$t-test^{*}$	$F-test^{**}$	$t-test^{*}$	$F-test^{**}$	$Model\ Fit^{***}$
Mining and Construction	✓	✓	×	✓	22.10
Nondurable Manufacturing (31)	✓	✓	✓	✓	5.90

Nondurable Manufacturing (32)	✓	✓	✓	✓	12.04
Durable Manufacturing	✓	✓	✓	✓	13.97
Trade and Transportation	✓	✓	✓	✓	2.50
Finance and Infrastructure	✓	✓	✓	✓	6.39
Health	✓	✓	✓	✓	3.04
Amusement	×	×	✓	✓	4.72
Repair	✓	×	✓	✓	14.23

* t-tests significant at 5% or less for one or more levels of the variable.
** Partial F-test significant at 5% or less.
*** If $\frac{Gener, Chi-Square}{DF}$ is close to 1 then the model is considered a good fit.

The findings for the relationship between magnitude of price changes and Region and Shipments and Receipts are these:

- All of the Regions were significant in explaining the magnitude of price changes in eight (8) of the nine (9) industry groups. Region was important in explaining the magnitude of price changes in seven (7) of the nine (9) industry groups as measured by the significance of F-tests. Region was not significant or important in explaining the magnitude of price changes in the Amusement industry group and it was not important in the Repair industry group.
- All of the levels of Shipments and Receipts were significant in explaining the magnitude of price changes in eight (8) of the nine (9) industry groups and was important in all nine (9) industry groups as measured by the F-test.

The general tendency observed is for the magnitude of price changes to be lower for smaller units than for larger units.

6 Estimating Nonresponse Bias

In this third and final phase of the nonresponse bias study, we used the results of our previous analyses to try to determine if nonresponse bias exists in PPI indexes. In the first phase of the nonresponse bias study we found a statistically significant association between a unit's Shipments and Receipts and response. In the second phase we found a statistically significant association between a unit's Shipments and Receipts and price movement. Though Region showed to be significant in the previous two phases, further study showed this variable to be highly correlated with industry, and since we sample by industry, we are unable to make adjustments for this variable.

In this phase we performed calculations to see if the differing levels of response by size class caused bias in our indexes. The calculations involved comparing one month percent changes based on our production, unadjusted, index values with those that had been calculated using nonresponse adjusted item weights. Statistically significant differences between these two index series would provide strong evidence to support the conclusion that bias exists in PPI indexes.

6.2 Data

With the exception of item weights, all of the data used to calculate both sets of indexes was final production data. In calculating the baseline index values production item weights were used. In calculating the non-response adjusted index values an adjustment factor was used to modify the production item weights. The derivation of this factor is described in the methodology section below.

We calculated original[7] and adjusted indexes for 272[8] 6-digit NAICS PPI industries for the period January, 2009 – December, 2009[9]. We calculated index percent changes for each month for each industry and subtracted the percent change for original indexes from the percent change for the adjusted indexes for each month. This gave us 12 months of index values and 11 months of percent changes for both the original and adjusted indexes. The analysis was performed on percent changes in the six digit net industry indexes.

6.3 Methodology

The item weights in the adjusted indexes were modified to account for total initiation and repricing non-response based on five size classes for each industry. [10] For each industry, we created a set of five non-response adjustment cells based on Shipments and Receipts quintiles. We calculated a Non-response Adjustment Factor (NRF) for each month based on the ratio of total responding and non-responding weight to responding weight. The NRF for each industry/S+R class/month represents the total productive and refusal sample weight divided by the responding weight. Since the factor can include item attrition without entire unit attrition, we used the sum of item weights for the responding portion but used the theoretically identical unit weights for the initiation portion because item weight is not available for non-responding units. The NRF was computed for each month using the following formula:

$$NRF_h = \frac{\sum_{s \in R} w_{hs} + \sum_{s \in N} w_{hs}}{\sum_{i \in R} w_{hi}},$$

where s = sample unit, i = item, and h = industry/quintile stratum. We adjusted the item weights in the following way: $w_{hi}^{'} = NRF_h * w_{hi}$

Non-response adjusted item weight = original item weight * NRF.

The Wilcoxon signed rank (WSR) test is a non-parametric[11] statistical hypothesis test used for testing whether the population medians from two related samples are equal. (The related samples may for example, be paired measurements on each of several

---

[7] To account for slight methodological changes in Janus over time we recalculated the unadjusted indexes we used in the comparison. While we refer to them as 'original' indexes they are really unadjusted indexes recalculated on the same platform as the adjusted indexes.

[8] The number of industries included in this phase of the study is less than in earlier phases. Phase 2 included 324 industries. To be included in the earlier phases industries had to have been sampled as a NAICS and have two years of history for the given sample prior to 2010. The two years of history could be any two year period. For this phase we needed to select a specific period to calculate indexes. We wanted to avoid sample changes within our calculation period. Consequently, 34 industries that switched samples during 2008 or 2009 had to be removed to conduct this analysis. 14 industries were removed because accurate non-response adjustment factors could not be calculated based on the collection data available within the size class cells. An additional four industries were removed because item weights had been manually updated and were inconsistent with the unit weights used in the NRF's numerator.

[9] We had intended to calculate index values for 2008 and 2009 but we had unforeseen issues using the Janus Estimation test environment and were only able to calculate 2009 in the time frame allotted.

[10] Item weights for nonresponse were adjusted on a month by month basis.

[11] This term is used to describe statistical techniques that do not rely on data belonging to any particular statistical distribution such as the normal distribution, for example.

subjects.) We take the differences among each pair, rank them, and use the ranks to compute the test statistic. (See Appendix I for further explanation of this procedure.)

6.4 Results

For the 272 industries, we calculated the following statistics: The minimum and maximum differences between the adjusted index percent change and the original index percent change, the Wilcoxon signed rank statistic for the differences between the two index series, and the p-value of the Wilcoxon signed rank statistic. Ten of these industries (3.7 percent) showed statistically significant signs of non-response bias at the $\alpha = 5\%$ level. Of these 10 industries, only 3 (326220 – Rubber and Plastics Hosing and Belting Manufacturing, 331316 – Aluminum Extruded Product Manufacturing, and 332439 – Other Metal Container Manufacturing) had maximum differences between the adjusted and original one month percent changes which could be considered large (0.3607, 0.4105 and 0.3457 respectively). The remaining 7 industries showing signs of non-response bias had negligible maximum differences between the adjusted and original indexes.

The industries that showed signs of non-response bias and had the largest differences between the original and adjusted percent changes were relatively unique in that they had experienced high percentages of non-response in one size class. The NRFs had reached 6.15, 7.46 and 8.67 for one adjustment cell in each of these industries. These NRFs were in the top two percent of all NRFs. The NRFs for the other adjustment cells within these industries did not show dramatic changes leaving the industries with a large range of NRFs. The large weight adjustment differential for items within the industry led to large changes in item relative importance between the original and non-response adjusted indexes. This led to the large difference in the index percent changes. These industries are also relatively old. Two had their current samples introduced in January 2005 and one had its current sample introduced in January 2006.

The effect on PPI indexes of adjusting item weights to account for initiation and repricing non-response was negligible. Very few industries exhibited nonresponse bias and those which did appear to be affected by unusually high nonresponse in very specific size classes. We conclude that PPI indexes as a whole do not suffer from nonresponse bias. The PPI is involved in ongoing efforts to improve response such as monitoring response more closely and in delinquency follow-up. We know that nonresponse will never be entirely eliminated especially in a voluntary survey, but with continued care in the selection of industries to be resampled and improvements to our sampling methodology, nonresponse (and nonresponse bias) may be kept at an acceptable minimum.

7 Conclusion

The objective of our research was to determine if nonresponse bias existed in PPI data. We found a statistically significant correlation between unit and item response and several frame variables including Employment, Region and Shipments and Receipts. Substituting Shipments and Receipts for Employment in modeling item prices, we found a statistically significant association between item prices and Region and Shipments and Receipts. Dropping Region from further consideration, we used Shipments and Receipts to create nonresponse adjustment cells and calculated item weights which were used in calculating adjusted indexes. Original and adjusted index estimates were calculated and compared to see if there were statistically significant differences between them. Very few industries exhibited signs of nonresponse bias. These findings did not identify strong evidence of nonresponse bias in PPI indexes for the industries and years analyzed.

8 Suggestions for Further Study

The focus of our analysis was on nonresponse bias in 6-digit NAICS industry indexes since that is how data are sampled and how any bias would have to be addressed. Data availability issues in Part 1 led us to limit the number of industries in our study to a total of approximately 300. One area of future research would be to increase the number of industries analyzed to more closely match the number for which the PPI calculates indexes. Another area of future research may be to investigate whether nonresponse bias exists in indexes calculated using other structures such as commodity and stage of processing indexes. This study compared adjusted and non-adjusted percent changes calculated for a one year period. Another possibility would be to calculate indexes for a longer time period.

References

Hollander, Myles and Wolfe, Douglas A. 1999. *Nonparametric Statistical Methods.* New York: John Wiley.

Conover, W.J. 1971. *Practical Nonparametric Statistics*. New York: John Wiley.

Agresti, Alan. 1996. *An Introduction to Categorical Data Analysis.* New York: John Wiley.

Hosmer, David W. and Lemeshow, Stanley. 2000. Applied Logistic Regression, Second Edition. New York: John Wiley.

Office of Management and Budget. 2006. "Standards and Guidelines for Statistical Surveys."

Verbeke, Geert, Spiessens, Bart and Lesaffre, Emmanuel. *Conditional Linear Mixed Model*. The American Statistician. February 2001. Vol. 55 No. 1.

Agresti, Alan, Booth, James G., Hobert, James P. and Caffo, Brian. 2009. *Random-Effects Modeling of Categorical Response Data*. Sociological Methodology. Vol. 30.

Hedeker, Donald. *Generalized Linear Mixed Models,* Encyclopedia of Statistics in Behavioral Science.

Schabenberger, Oliver. 2004. *Mixed Model Influence Diagnostics*, SAS Institute Inc; SAS Institute Inc. *SAS/STAT® user's guide, version 9.1*. Cary, NC: SAS Institute Inc.

Federal Committee on Statistical Methodology (FCSM). 2001. *Measuring and Reporting Sources of Error in Surveys.* Statistical Policy Working Paper 31, Chapter 4.

Collett, D. 2003. *Modeling Survival Data in Medical Research*. Chapman & Hall.

Allison, Paul D. 1995. *Survival Analysis Using SAS: A Practical Guide*. SAS Publishing.