SUPPORTING STATEMENT
REGIONAL ECONOMIC DATA COLLECTION PROGRAM
FOR SOUTHWEST ALASKA
OMB CONTROL NO.: 0648-xxxx

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used. Data on the number of entities (e.g. establishments, State and local governmental units, households, or persons) in the universe and the corresponding sample are to be provided in tabular form. The tabulation must also include expected response rates for the collection as a whole. If the collection has been conducted before, provide the actual response rate achieved.

For the vessel surveys, the overall population consists of all fishing vessels landing raw fish at a port in Southwest, Alaska during 2005. For that year, there were 2,117 vessels. This population consists of three vessel classes – small, medium, and large vessel classes. The population sizes are 1,479, 421, and 217, respectively for small, medium, and large vessels classes. An unequal probability sampling (UPS) procedure is used to determine the sample sizes needed for the analysis for each vessel class, which is described in Item #2 below and in Attachment D. The population sizes of local businesses and fish processors are 172 and 41, respectively.

The expected response rates for the vessel surveys are based on consideration of the following factors. First, compared with a previous data collection project conducted for Southeast Alaska (Hartman 2002), which achieved an overall response rate of about 30%, the number of questions in the present project is much smaller and the quantity of information being asked is much smaller. Second, in the present study, questions about sensitive information such as vessel cost and expenditures are omitted. The previous Southeast study included these sensitive questions, which significantly contributed to the low response rate. Third, input from select members of the respondent populations helped guide survey design and question wording. Fourth, follow-up telephone calls will also increase the response rate. Based on these factors, it is expected that, overall, the response rate for mail survey of fishermen for the present project will be about 55% which is much higher than in the Southeast study. For telephone interviews with local businesses (including fish processors), a response rate of 65% is expected. For a more detailed description of the methods we used, and will use, to increase the response rate, see Item #3 below.

Vessel Class	Population size	Mail or phone interview sample size	Expected number of respondents	Expected response rate
Small vessel	1,479	491	270	55%
Medium vessel	421	225	124	55%
Large vessel	217	164	90	55%
Local businesses including fish processors	213	213	139	65%

2. Describe the procedures for the collection, including: the statistical methodology for stratification and sample selection; the estimation procedure; the degree of accuracy needed for the purpose described in the justification; any unusual problems requiring specialized sampling procedures; and any use of periodic (less frequent than annual) data collection cycles to reduce burden.

Since the majority of gross revenue within each harvesting sector comes from a small number of vessels, a simple random sampling (SRS) of vessels would only include a small portion of the total ex-vessel value, and therefore, would be misleading. As a result, for the present project an unequal probability sampling (UPS) method without replacement is used that accounts for this unequal harvest in each target population. The objective of implementing the sampling task is to estimate the employment and labor income information for each of three disaggregated harvesting sectors using as an auxiliary variable the ex-vessel revenues provided by Commercial Fisheries Entry Commission (CFEC) earnings data. Since each sector will be used as a separate economic sector in IMPLAN model, we face three separate problems for three different sectors in sampling. For each sector, we use a UPS without replacement method to identify sampling units. Details on our sampling methodology are described in Attachment D.

3. Describe the methods used to maximize response rates and to deal with non-response. The accuracy and reliability of the information collected must be shown to be adequate for the intended uses. For collections based on sampling, a special justification must be provided if they will not yield "reliable" data that can be generalized to the universe studied.

(a) Maximizing Response Rates

Previous applications of voluntary commercial fishing surveys in Alaska (e.g., Hartman 2002) were hampered by low response rates that principally resulted from the use of long and complicated survey instruments. Commercial fishermen are frequently asked, and often required, to participate in surveys from numerous organizations including NOAA, Alaska Department of Fish and Game (ADF&G), and universities. As a result, commercial fishermen are less likely to complete voluntary surveys that are lengthy, poorly-designed, and do not clearly involve issues that are important to them. In this data collection, significant efforts were made to ensure the survey instruments were short in length, contained well-designed questions, and clearly conveyed the importance of the data collection to issues that are important to commercial fishermen.

The mail surveys are short (6 to 7 questions depending upon the survey version, all of which span eight pages) and avoid many sensitive questions compared with many previously-fielded commercial fishing surveys. The set of questions was limited to only those that are essential for achieving the objectives of the project as outlined in Part A, Item #1 above. There is only a fraction of the number of questions asked compared with the Southeast Alaska commercial fishing survey discussed earlier, which achieved an overall response rate of about 30%. In the mail surveys, numerous questions on vessel expenditures that are often included in surveys of commercial fishermen are omitted here to avoid the added complexity and likely sensitivity of asking for this type of information from respondents.[1]

[1] Vessel expenditures will be estimated using (1) the sales data collected from telephone interviews with local

The telephone scripts for use in interviews with local businesses and fish processors were developed with similar goals in mind. Specifically, each phone script was constructed to include only the most essential questions to ensure the telephone interviews were short in length to minimize the time burden on respondents.

Pretesting activities that included a small focus group and several interviews with fishermen and fish processors (totaling less than 10 individuals) were used to evaluate the content and presentation of the survey materials, as well as to ensure input by the fishing community. Feedback from these pretesting activities aided in non-trivial ways to the development of the survey questions. For instance, considerable effort was made to ensure that the survey instrument reflected considerations for the record-keeping systems kept by fishermen and used common terms and wording used by fishermen. Participants in pretesting activities also indicated that previous voluntary surveys often did not provide adequate assurances that the information being requested would be handled confidentially, which often deterred them from responding. To ensure respondents that the data they share will be kept confidential, a detailed confidentiality statement is presented on the first page of the mail survey and mentioned upfront in the telephone interviews. A similar statement is made in the cover letter accompanying the mail survey.

Another reason believed to have caused low response rates in previous survey efforts is the disinterest among respondents toward the survey purpose. Surveys that collect information that will clearly benefit or interest respondents are more likely to be completed. The importance and benefits of this data collection project to the respondents (fishermen, local businesses, and fish processors) will be emphasized in the advance letter, cover letter, mail survey, and telephone interviews. In these letters and phone interviews, the investigators clearly state that with the help of the respondents, the important role of the respondents' fishing and business activities in the regional economy can be better identified and that the information they provide will be used to enhance the fishery management practices of NOAA Fisheries, and, thereby, to increase the long-run economic benefits to the fishermen and local businesses. Making a clear link between the survey, their participation, and the fishery and regional economy is expected to help increase the response rate relative to previous studies.

In addition to the above steps taken to maximize response rates, the survey instruments (mail and telephone) were subjected to significant review by several researchers with expertise on Alaska fisheries and economic surveys to ensure the quality of the materials.

In addition to high-quality survey instruments, the set of survey protocols to be followed in implementation was designed to maximize response rates. For the mail survey, a modified Dillman (2000) approach will be employed that includes four survey contacts as follows (All the letters, postcard reminder, and follow-up phone scripts for these four contacts are attached in Attachment C):

- An **advance letter** notifying the respondents a few days before they receive the survey questionnaire. This will be the first contact with the respondent.
- An **initial mailing** sent a few days after the advance letter. Each mailing will contain

---

businesses and fish processors and (2) a cost engineering approach.

a cover letter, personalized questionnaire, and a pre-addressed stamped return envelope.

- A **postcard follow-up reminder** mailed 5-7 days following the initial mailing.
- A **follow-up phone call** to encourage response and identify individuals that have misplaced or need another copy of the survey. If the respondent agrees, the mail survey will be completed over the phone.[2] Up to three attempts will be made to contact each respondent for the telephone interview. Individuals needing an additional copy of the survey will be sent one with another cover letter and return envelope.

A strict Dillman approach is not warranted, given negative input from commercial fishermen about repeated contacts beyond the phone contact.

The result of the efforts described above are compact and high-quality survey instruments that contain questions vessel owners, local businesses, and fish processors can answer with minimal effort. As a result, the expected response rate for the mail survey of fishermen is expected to exceed previous survey efforts and achieve a response rate of approximately 55%. This response rate is much higher than that in the longer and more complicated Southeast Alaska study (30% response rate). For the telephone interviews with local businesses (including fish processors), a response rate of 65% is assumed based on previous experience.[3]

(b) Non-response

To better understand the differences between them, comparisons will be drawn between respondents and non-respondents with respect to several observable characteristics: (1) geographical area of landed fish, (2) ex-vessel value, and (3) species that vessels catch. This information is available from government data for each vessel. If significant and systematic differences between the two groups are discovered, the population parameter estimates of interest may be adjusted by using weights formed from these variables.

4. Describe any tests of procedures or methods to be undertaken. Tests are encouraged as effective means to refine collections, but if ten or more test respondents are involved OMB must give prior approval.

There are no plans to conduct a pilot survey or other tests involving more than ten respondents.

5. Provide the name and telephone number of individuals consulted on the statistical aspects of the design, and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

John Slanta (Census Bureau, PH 301-763-4773) and Dr. Dan Lew (NMFS, PH 206-526-4252) assisted in the development and review of sampling procedures for this project.

---

[2] In this case, the ex-vessel values (by species) of the vessel will be provided to the vessel owners so that they will not have to access their records, which should greatly simplify the question and allow them to calculate the crew and skipper payments easily. In doing this, we will make sure that the person we will be interviewing on the phone is the true owner of the vessel. This is because we do not want to breach the confidentiality by providing the sensitive information to the wrong person. As is seen in the mail survey questions (Attachment A), however, this ex-vessel information will not be given to the respondent in the mail survey.

[3] See Section A #12, Footnote 6.

Several NMFS economists with experience in economic survey design and implementation reviewed the survey materials and survey protocols, including Dr. Dan Lew, Dr. Ron Felthoven, and Dr. Brian Garber-Yonts.

Professor Hans Geier (University of Alaska, Fairbanks) is the contractor who will conduct the data collection project, revise the IMPLAN data, and participate in developing regional economic models.

Dr. Chang Seung (Alaska Fisheries Science Center) will conduct the statistical analysis of the information collected, and develop regional economic models with Professor Geier.

ATTACHMENT A. SAMPLING PROCEDURES FOR HARVESTING SECTORS[1]

The overall project objective is to estimate the employment and labor income information for each of three disaggregated harvesting sectors using data to be collected via a mail survey. Using ex-vessel revenue information, an unequal probability sampling (UPS) procedure will be employed to determine the sampling plan for each of the three harvesting sectors. The procedure is described below.

In the literature, there exist many methods for conducting UPS without replacement (see, for example, Brewer and Hanif 1983; Sarndal 1992). One critical weakness with most of these methods is that the variance estimation is very difficult because the structure of the 2nd order inclusion probabilities (π_{ij})[2] is complicated. One method that overcomes this problem is Poisson sampling. However, one problem with Poisson sampling is that the sample size is a random variable, which increases the variability of the estimates produced. An alternative method that is similar to Poisson sampling but overcomes the weakness of the Poisson sampling is Pareto sampling (Rosen 1997)[3] which yields a fixed sample size.

In this project, there are two tasks that we need to do for estimating the population parameters using UPS without replacement. First, the optimal sample size needs to be determined. Second, once the optimal sample size is determined, the population parameters and confidence intervals need to be estimated. For the first task, we will use the variance of Horvitz-Thompson (HT) estimator from Poisson sampling in Part I below.[4] For the second task, we will use the Pareto sampling method described in Part II below (Slanta 2006). In determining the optimal sample size in Part I, we will use information on an auxiliary variable (ex-vessel revenue). To estimate the population parameters in Part II, we use actual response sample information on the variables of interest (employment and labor income).

Part I: Estimating Sample Size

Step 1: Estimation of Optimal Sample Size (n*)

(A) Obtaining Initial Probabilities

To obtain the initial values of the inclusion probabilities (π_i) for unit i in the population, we multiply the auxiliary value of unit i (X_i, i.e., the ex-vessel value of vessel i in the population) by a proportionality constant (t)[5]:

$$\pi_i = t X_i \qquad (1)$$

where π_i : probability of vessel i being included in the survey sample

X_i : value of the auxiliary variable (ex-vessel value of vessel i in the population)

Here, t is given by

$$t = \frac{\sum_{i}^{N} X_i}{V + \sum_{i}^{N} X_i^{\,2}} \tag{2}$$

where N : population size

V : desired variance (of HT estimator of the population total); Poisson variance. Here, V is given as:

$$V = \left(\frac{\varepsilon X}{z_{1-(\alpha/2)}} \right)^2$$

where ε is the error allowed by the investigator [e.g., if ε is 0.1, then 10% error of true population total ($X = \sum_{i=1}^{N} X_i$) is allowed]; and z is percentile of the standard normal distribution. Therefore, choosing a desired variance V is equivalent to setting the values of ε and z. The value of V calculated using $V = \sum_{i=1}^{N} \frac{(1-\pi_i)X_i^2}{\pi_i}$ (Poisson variance; Brewer and Hanif 1983, page 82) with π_i's being the final values of N inclusion probabilities obtained from Step 1, will be equal to the desired variance given at the beginning of Step 1.

Some of the resulting π_i's could be larger than one. The number of certainty units (i.e., the number of units for which $\pi_i > 1$) is denoted C_1. If $\pi_i > 1$, then we force this inclusion probability to equal one ($\pi_i = 1$).

(B) Iterations and Determination of Optimal Sample Size

We recalculate t using the noncertainty units (i.e., the units for which $\pi_i < 1$) obtained in (A) above, i.e.,

$$t = \frac{\sum_{i}^{M_1} X_i}{V + \sum_{i}^{M_1} X_i^{\,2}} \tag{2'}$$

where M_1 : number of noncertainty units from (A), where $M_1 = N - C_1$.

Using equation (1) above, we calculate the inclusion probabilities for the noncertainty units by multiplying the t value [from equation (2')] by the ex-vessel values of the noncertainty units. If the resulting π_i's are larger than one, we force them to equal one. The resulting numbers of certainty and noncertainty units are denoted C_2 ($= C_1$ + additional number of certainty units) and M_2 ($= M_1$ – additional number of certainty units), respectively, where $C_2 + M_2 = N$. Next, for M_2 units of noncertainty, we calculate the t and π_i's again. This is an iterative process. We continue this process until the noncertainty population stabilizes (i.e., until there is no additional certainty unit).

If the noncertainty population stabilizes after *k*th iteration, there will be C_k units of certainty units and M_k units of noncertainty units and $C_k + M_k = N$. Summing over the probabilities for all these certainty and noncertainty units, we obtain the optimal sample size (n*) as:

$$n^* = \sum_{i}^{N} \pi_i \qquad (3)$$

At this stage the optimal sample size may not be an integer number. In this stage, we also compute the optimal sample size under simple random sampling (SRS)[6], n_{srs}, and compare it with n*.

<u>Step 2: Determining Number of Mailout Surveys</u>

(A) Adjustment of Probabilities

Once the optimal sample size (n*) is determined in Step 1, we divide the sample size (n*) by the expected response rate (obtained from previous studies) to determine the number of surveys that need to be mailed out to achieve n*. The number thus derived is denoted n_a (this number may not still be an integer value). We next adjust the inclusion probabilities for the M_k noncertainty units obtained in Step 1 above as:

$$\pi_i = (n_a - C_k)\left[\frac{\pi_i}{\sum_{i}^{M_k} \pi_i}\right] \qquad (4)$$

If the resulting probabilities are larger than one ($\pi_i > 1$), we make them certainties ($\pi_i = 1$). The resulting numbers of certainty and noncertainty units are denoted C_{k+1} and M_{k+1}, respectively. Next, we adjust the probabilities of the new set of noncertainty units (M_{k+1}) in a similar way using equation (4') below:

$$\pi_i = (n_a - C_{k+1})\left[\frac{\pi_i}{\sum_{i}^{M_{k+1}} \pi_i}\right] \qquad (4')$$

We continue this process until the noncertainty population stabilizes. The resulting numbers of certainty and noncertainty units are C_q and M_q, respectively.

(B) Apply Minimum Probability Rule

At this point, we impose a minimum probability rule. UPS can have excessively large weights ($= 1/\pi_i$) and if they report a large value, then the population estimate and its variance would be very large. In order to avoid this problem, we can impose a minimum value of the inclusion probabilities. If *m* is the minimum imposed probability, then we do the following:

If $\pi_i < m$, then set $\pi_i = m$ for each i, where i = 1, ..., N.

The value for *m* here is determined arbitrarily. The only cost involved in using this rule is a small increase in sample size.[7]

(C) Finding an Integer Value for Sample Size

Next, we add up all the resulting inclusion probabilities. The resulting sum is denoted n_b (> n_a), which may not be an integer value. Next, we adjust again the probabilities for noncertainty units including the units for which the minimum probabilities were imposed as:

$$\pi_i = (n_c - C_q)\left[\frac{\pi_i}{\sum_i^{M_q} \pi_i}\right] \qquad (5)$$

where n_c is the smallest integer value larger than n_b (e.g., if $n_b = 15.3$, then $n_c = 16$). Finally, we add up the resulting (certainty and noncertainty) probabilities. The sum of all these probabilities is the final survey sample size (i.e., the number of surveys to be sent out to), and is denoted n_m (= n_c).

Part II: Estimation of Population Parameters and Confidence Intervals

Step 3: Implementation of Pareto Sampling

After the mailout sample size (n_m) for each sector is determined in Step 2, the mailout sample is selected from each sector's population using Pareto sampling. The probability of each unit (vessel) being in the sample in a given sector is proportional to the unit's (vessel's) ex-vessel revenue. Because the majority of gross revenue within each sector comes from a small number of vessels, a random sample of vessels would only include a small portion of the total ex-vessel values.

According to Brewer and Hanif (1983), there are fifty different approaches that are used for UPS. Most of these approaches suffer from the weakness that it is very hard to estimate the variance. Poisson sampling overcomes this problem, and is relatively easy to implement. However, the limitation of Poisson sampling is that the sample size is a random variable. Therefore, in this project, we will use Pareto sampling (Rosen 1997 and Saavedra 1995) which overcomes the limitation of Poisson sampling. The mailout sample size will be n_m as determined in Step 2 (C) above. We will use the inclusion probabilities obtained from Equation (5) above in implementing Pareto sampling.

The procedure of this sampling method (Block and Crowe 2001) is briefly described here:

1. Determine the probability of selection (π_i) for each unit i as in Equation (5) above.

2. Generate a Uniform (0,1) random variable U_i for each unit i
3. Calculate $Q_i = U_i (1 - \pi_i) / [\pi_i (1 - U_i)]$
4. Sort units in ascending order by Q_i, and select n_m smallest ones in sample.

From the above, it is clear that we will have a fixed sample size with Pareto sampling.

Step 4: Mailing out Surveys and Obtaining Actual Response Sample

Next, we will send out the surveys to the n_m units (vessel owners). Actual response sample will be obtained and the size of the actual response sample is denoted r.

Step 5: Estimation of Population Parameters (Population Total)

Using the information in the actual response sample, we calculate population parameters *for variables of interest* (employment and labor income in our project), *not for ex-vessel revenue,* using HT estimator (Horvitz and Thompson 1952). We are interested in estimating the population totals (not population means) of the variables of interest. The HT estimator is given as:

$$\hat{Y}_{HT} = \sum_{i=1}^{r} w_i y_i \qquad (6)$$

where r : number of respondents
w_i : weight for ith unit (= $1/\pi_i$). Note that the weights are calculated here using the information on the auxiliary variable, not that on the variables of interest
y_i : response sample data of i^{th} unit (employment or labor income)

However, the HT estimator needs to be adjusted for non-response. The estimator is adjusted in the following way.

$$\hat{Y} = \left(\frac{\sum_{j=1}^{N} X_j}{\sum_{i=1}^{r} w_i X_i} \right) \hat{Y}_{HT} \qquad (7)$$

where N : population size
X_i : auxiliary variable of i^{th} unit (respondents only)

Usually, we apply this adjustment to the certainties separately from the noncertainties, and then add the two together to get a final estimate. If there are no respondents within any of the two groups of certainty units and noncertainty units, then we collapse the two groups before applying the adjustment. Specifically, the final estimate of population total is given by:

$$\hat{Y} = \left(\frac{\sum_{j=1}^{N_1} X_j}{\sum_{i=1}^{r_1} w_i X_i} \right) \sum_{i=1}^{r_1} w_i y_i \quad + \quad \left(\frac{\sum_{j=1}^{N_2} X_j}{\sum_{i=1}^{r_2} w_i X_i} \right) \sum_{i=1}^{r_2} w_i y_i \tag{8}$$

where N_1 : number of certainty units in the population
N_2 : number of noncertainty units in the population
r_1 : number of respondents from certainty units
r_2 : number of respondents from noncertainty units, and
$N_1 + N_2 = N$ and $r_1 + r_2 = r$.

Step 6: Estimation of Variance for $\hat{Y}_{HT}$ and $\hat{Y}$

Here we will calculate the variances of the population estimates for the variables of interest. The variance estimate for Pareto sampling is given in Rosen (1997, Equation (4-11), p. 173) as:

$$Var(\hat{Y}_{HT}) = \frac{n_m}{n_m - 1} \left\{ \left[\sum_{i=1}^{n_m} (1-\pi_i) \left(\frac{y_i}{\pi_i} \right)^2 \right] - \frac{\left[\sum_{i=1}^{n_m} y_i \left(\frac{1-\pi_i}{\pi_i} \right) \right]^2}{\sum_{i=1}^{n_m} (1-\pi_i)} \right\} \tag{9}$$

Since we have adjusted for nonresponse, we need to incorporate the variability due to nonresponse into the variance. If we assume that the response mechanism is fixed [8], then we have a ratio estimator and its variance can be found in Hansen, Hurwitz, and Madow (1953, page 514). This variance is a Taylor expansion, and is given as:

$$Var(\hat{Y}) = \hat{Y}^2 \left(\frac{\hat{\sigma}^2(A)}{A^2} + \frac{\hat{\sigma}^2(B)}{B^2} - \frac{2COV(A,B)}{AB} \right) \tag{10}$$

where

$$A = \sum_{i=1}^{r} w_i y_i$$

$$B = \sum_{i=1}^{r} w_i X_i$$

$$\hat{\sigma}^2(A) = \frac{n_m}{n_m - 1} \left\{ \left[\sum_{i=1}^{r} (1-\pi_i)(w_i y_i)^2 \right] - \frac{\left[\sum_{i=1}^{r} (1-\pi_i)(w_i y_i) \right]^2}{\sum_{i=1}^{n_m} (1-\pi_i)} \right\}$$

$$\hat{\sigma}^2(B)=\frac{n_m}{n_m-1}\left\{\left[\sum_{i=1}^{r}(1-\pi_i)(w_iX_i)^2\right]-\frac{\left[\sum_{i=1}^{r}(1-\pi_i)(w_iX_i)\right]^2}{\sum_{i=1}^{n_m}(1-\pi_i)}\right\}$$

$$COV(A,B)=\frac{n_m}{n_m-1}\left\{\left[\sum_{i=1}^{r}(1-\pi_i)w_i^2y_iX_i\right]-\frac{\left[\sum_{i=1}^{r}(1-\pi_i)(w_iy_i)\right]\left[\sum_{i=1}^{r}(1-\pi_i)(w_iX_i)\right]}{\sum_{i=1}^{n_m}(1-\pi_i)}\right\}.$$

Step 7: Calculation of Confidence Intervals

Confidence intervals are calculated using response sample statistics obtained in steps 5 and 6. We only choose one sample, but if there were many independent samples chosen then we would expect on average that approximately 100(1-α) % of the confidence intervals constructed in the following manner will contain the truth.

$$\left(\hat{Y}-z_{\alpha/2}\sqrt{Var(\hat{Y})}\ ,\ \hat{Y}+z_{\alpha/2}\sqrt{Var(\hat{Y})}\right) \qquad (11)$$

where $\hat{Y}$: Estimated population total for employment or labor income.

Note that it is possible to use t-statistics if the sample size is small.

Footnotes

1. In the process of developing this document, several experts in UPS sampling assisted me by providing helpful comments and inputs. The experts include John Slanta (U.S. Census Bureau), Bengt Rosen (Uppsala University), Pedro Saavedra (ORC Macro), Holmberg Anders (Statistics Sweden), Paolo Righi (ISTAT, Italy), and Bob Fay (U.S. Census). In particular, I would like to thank John Slanta very much for his time and effort in providing valuable inputs and advice. His suggestions and comments contributed significantly to the development of the sampling procedures in this document. Many thanks go to Dan Lew (NMFS) for his rigorous review and valuable suggestions which contributed in a significant way to the improvement of this document. I also benefited from discussions of UPS with Norma Sands at NWFSC and from the Excel file that she developed.

2. 2nd order inclusion probability (π_{ij}) is defined as the joint probability of including in sample the ith and jth population units.

3. Saavedra (1995) independently developed the same sampling methodology as Rosen (1997), which he called Odds Ratio Sequential Poisson Sampling (ORSPS).

4. Although we do not use Poisson sampling itself, we do use the Poisson variance of HT estimator of the population total.

5. Equation (1) is derived as follows.

   HT estimator, $\hat{X}_{HT} = \sum_i \frac{X_i}{\pi_i}$, has variance,

   $$V(\hat{X}_{HT}) = \sum_{i=1}^{N} \frac{X_i^2}{\pi_i}(1-\pi_i) = \sum_{i=1}^{N} \frac{X_i^2}{\pi_i} - \sum_{i=1}^{N} X_i^2 \quad \text{(Brewer and Hanif 1983, page 82)} \quad \text{(A)}$$

   For an expected sample size *n*,

   $$\pi_i = n\left(\frac{X_i}{\sum_{i=1}^{N} X_i}\right) \quad \text{(B)}$$

   Substituting (B) into (A) and solving for n,

   $$n = \left(\sum_{i=1}^{N} X_i\right)^2 \Big/ \left(V(\hat{X}_{HT}) + \sum_{i=1}^{N} X_i^2\right) \quad \text{(C)}$$

   Substituting (C) into (B),

   $$\pi_i = \left[\frac{\sum_{i=1}^{N} X_i}{V(\hat{X}_{HT}) + \sum_{i=1}^{N} X_i^2}\right] X_i, \quad i = 1, 2, \ldots, N, \quad \text{(D)}$$

where $V(\hat{X}_{HT})$ is the desired variance.

6. The optimal sample size under SRS is determined using the following standard formula:

$$n_{srs} \geq \frac{z^2 N (CV_p)^2}{z^2 (CV_p)^2 + (N-1)\varepsilon^2}$$ (Levy and Lemeshow, formula (3.14) on page 74)

where n_{srs} : optimal sample size under SRS

CV_p : coefficient of variation of the population parameter. Since the information on the population parameters (i.e., employment and labor income) is not available, we use ex-vessel revenue, for which the population information is available from CFEC. Therefore, CV_p is defined as standard deviation of the ex-vessel revenue in the population divided by the mean.

7. This minimum probability rule is used, for example, in the Manufacturing and Construction Division of the Census Bureau. To date, there has not been any research on the minimum probability in the sampling literature. It is an arbitrary value and in applications has sometimes varied between strata in the same survey. Some researchers determine the minimum probability such that the resulting weight, which is the reciprocal of the minimum probability, is less than or equal to the population size. Generally speaking, this minimum probability rule has little effect on the sample size.

8. Fixed response mechanism means that a unit included in a sample is always a respondent or non-respondent no matter what sample the unit is included in. In other words, the probability of the unit being a respondent is either one or zero but nothing in-between.

References

Block, C. and Crowe, S. (2001). Pareto-πps Sampling. Unpublished Document. Statistics Canada.

Brewer, K. and Hanif, M. (1983). *Sampling with Unequal Probabilities*. Springer Verlag, New York.

Hansen, Hurwitz, and Madow (1953). *Sampling Survey Methods and Theory*. Volume 1. Methods and Applications.

Horvitz, D. and Thompson, D. (1952). A Generalization of Sampling without replacement from a Finite Universe. Journal of American Statistical Association Vol. 47, pp. 663-685.

Levy, P. and Lemeshow, S. (1999). *Sampling of Populations – Methods and Applications.* Third Edition. Wiley and Sons.

Rosén, B. (1997). On Sampling with Probability Proportional to Size. *Journal of Statistical Planning and Inference*, 62, 159-191.

Särndal, C.-E., Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer Verlag, New York.

Saavedra, P. 1995. Fixed Sample Size PPS Approximations with a Permanent Random Number. Joint Statistical Meetings, American Statistical Association, Orlando, Florida.

Slanta, J. (2006). Personal Communication.