Sample Design and Weighting Procedures for the BiH STEP Employer Survey

David J. Megill
Sampling Consultant, World Bank
May 2017

1. Sample Design for BiH STEP Employer Survey

The sampling frame for the Bosnia and Herzegovina (BiH) STEP Employer Survey was based on the business register of all enterprises in Bosnia and Herzegovina with 5 or more employees, with information on the geographic location, the economic activity and the number of employees. The enterprises in the sampling frame were stratified by three regions (Sarajevo, Rest of Federation and Republic of Srpska) and three size groups in terms of the number of employees (5-19, 20-99 and 100+). A stratified two-stage sample design was used for the BiH STEP Employer Survey, with a sample of enterprises selected within each stratum at the first stage, and branches selected at the second stage. Table 1 presents the distribution of the firms (enterprises) in the frame by region and employment size strata, and Table 2 shows the corresponding distribution of the total number of employees in each stratum.

Table 1. Distribution of firms (enterprises) in the sampling frame for the BiH STEP Employer Survey by region and employment size strata

Region	5-19 employees	20-99 employees	100+ employees	Total
(1) Sarajevo	1,442	463	125	2,030
(2) Rest of Federation	3,662	1,214	270	5,146
(3) Republic of Srpska	2,308	848	218	3,374
Total	7,412	2,525	613	10,550

Table 2. Distribution of total number of employees in the firms in the sampling frame for the BiH STEP Employer Survey by region and employment size strata

Region	5-19 employees	20-99 employees	100+ employees	Total
(1) Sarajevo	12,917	18,979	48,058	79,954
(2) Rest of Federation	33,302	48,566	90,414	172,282
(3) Republic of Srpska	21,598	34,491	70,996	127,085
Total	67,817	102,036	209,468	379,321

First it was necessary to allocate the sample by region and employment size strata based on the distribution of the frame and the sample size needed for each domain. Given that each region and employment size stratum was a domain of analysis, the sample of 504 enterprises was allocated equally to all the strata (with 56 sample firms each), as shown in Table 3. Given the distribution of the frame shown in Table 1, this resulted in a higher sampling rate for the larger employment size strata. A supplemental sample of up to 200% of the target sample size was also selected in each stratum as a reserve for possible replacements. In the case of strata that did not have triple the number of sample firms in the frame, all of the sample enterprises were selected in the initial phase.

Table 3.   Allocation of sample firms (enterprises) for BiH Employer Survey by region and employment size stratum

Region	5-19 employees	20-99 employees	100+ employees	Total
(1) Sarajevo	56	56	56	168
(2) Rest of Federation	56	56	56	168
(3) Republic of Srpska	56	56	56	168
Total	168	168	168	504

Within each region by employment size stratum the enterprises were selected systematically with probability proportional to size (PPS), after sorting the frame geographically and by economic activity. The measure of size in this case was the number of employees for each enterprise in the frame.

In the case of enterprises that have more than one branch (location), one branch was selected with equal probability at the second stage, except for large establishments that had been selected with a probability of 1 at the first stage. The number of branches is generally correlated with the number of employees, so this sampling strategy should reduce the variability in the weights. In the case of large enterprises that have a measure of size greater than the sampling interval, the number of branches to be selected was determined based on the number of "hits", as explained later.

The enterprises were selected at the first stage systematically with PPS within each region by employment size stratum, using the number of employees as the measure of size. The following steps were used for this sample selection:

1. Since the sample enterprises are selected systematically with PPS within each stratum, it is first necessary to calculate the sampling interval, which is equal to the cumulated total number of employees in the frame for the stratum divided by the number of enterprises to be selected in the stratum.

2. Any enterprise with a measure of size (number of employees) greater than the sampling interval for the stratum was selected with certainty (that is, with a probability of 1). These

2

self-representing (SR) enterprises were separated from frame and included in the original sample to be interviewed.

3. For each SR enterprise, divide the number of employees in the frame by the original sampling interval, and round to the next integer (for example, 1.5 would be rounded to 2) in order to determine the number of "hits". The number of "hits" will correspond to the number of branches to be selected in the SR enterprise.

4. Sum the number of "hits" for all SR enterprises in the stratum. Subtract this number from the total number of sample branches allocated to the stratum to determine the number of non-self-representing (NSR) sample enterprises to be selected in the stratum.

5. Select the NSR sample enterprises in each stratum systematically with PPS after separating the SR enterprises from the frame. It will be necessary to cumulate the measures of size (number of employees) again (excluding the SR enterprises) and calculate a new sampling interval.

6. After each iteration of the systematic PPS selection, it is necessary to separate the SR sample firms (those with a measure of size greater than the sampling interval), and adjust the number of NSR sample firms to be selected accordingly. Then a new sampling interval will be calculated for the selection of the NSR sample firms.

7. For all the sample SR and NSR sample enterprises, it will be necessary to make a list of all the workplaces (branches) in each.

8. At the second stage, select one branch from each sample NSR enterprise with equal probability. The headquarters should be counted as a branch and be listed as one of the possible branches to be selected for interviewing. For each SR enterprise, the number of "hits" will determine the number of sample branches to be selected with equal probability.

Once the original sample of SR and NSR enterprises had been selected for each stratum, it was necessary to contact them to obtain a list of their branches and then randomly select a branch to interview. The total number of branches in each sample enterprise was recorded since this information was needed for the calculation of the weights.

Following the implementation of the BiH STEP Employer Survey, a total of 536 sample branches were interviewed. In some cases after a sample firm was replaced, it was possible to complete the interview in the original firm as well as the replacement, so the target sample size was exceeded for most strata. The distribution of the final sample of branches by region and employment size stratum is shown in Table 4.

Table 4.   Final distribution of sample branches with completed interviews for BiH Employer Survey by region and employment size stratum

Region	5-19 employees	20-99 employees	100+ employees	Total
(1) Sarajevo	63	58	57	178
(2) Rest of Federation	67	58	63	188
(3) Republic of Srpska	58	58	54	170
Total	188	174	174	536

2. Weighting Procedures for BiH STEP Employer Survey

In order for the sample estimates from the BiH STEP Employer Survey data to be representative of the population of enterprises, it is necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample branch is equal to the inverse of its probability of selection.

As described above, a stratified two-stage sample design was used for the BiH STEP Employer Survey. At the first stage a sample of enterprises was selected in each region by employment size stratum systematically with PPS, based on the number of employees in the frame. At this stage some of the enterprises were selected with a probability of 1 because of their size, so they are considered to be self-representing (SR), that is, selected with certainty at the first stage. At the second stage more than one branch can be selected from the large SR enterprises with a measure of size that is a multiple of the sampling interval. In the case of the non-self-representing (NSR) sample enterprises, only one branch is selected in each enterprise at the second stage. The weights are specified here separately for the SR and NSR sample enterprises.

For the SR enterprises, the probabilities of selection can be expressed as follows:

$$p_{Shi} = \frac{b_{hi}}{B_{hi}},$$

where:

p_{Shi} = probability of selection for the sample branches in the i-th SR enterprise in stratum (region, employment size) h

b_{hi} = number of sample branches interviewed for the i-th SR enterprise in stratum h

B_{hi} = total number of branches identified in the frame for the i-th SR enterprise in stratum h

In this case the first stage probability of selection is 1, so it does not appear in the formula for the overall probability of selection. The basic weight for the SR sample enterprises is the inverse of this probability of selection, and can be expressed as follows:

$$W_{Shi} = \frac{B_{hi}}{b_{hi}},$$

where:

W_{Shi} = basic weight for the sample branches in the i-th SR enterprise in stratum h

For the NSR sample enterprises, the overall probabilities of selection within each stratum includes components from the first and second sampling stages. This probability can be expressed as follows:

$$p_{Nhi} = \frac{n'_h \times E_{hi}}{E_{Nh}} \times \frac{1}{B_{hi}},$$

where:

p_{Nhi} = probability of selection for the sample branch in the i-th sample NSR enterprise in stratum h

n'_h = number of NSR sample enterprises in stratum h with completed interviews, including any replacements that were interviewed

E_{hi} = number of employees in the frame for the i-th NSR enterprise in stratum h

E_{Nh} = total number of employees in the frame for all the NSR enterprises in stratum h (that is, the cumulated measure of size)

B_{hi} = total number of branches identified in the frame for the i-th sample NSR enterprise in stratum h

The two components of this probability correspond to the individual sampling stages. The first stage probability is based on the selection of the sample NSR enterprises with PPS within each stratum, based on the number of employees in the frame. By using the final number of NSR sample firms that are interviewed in each stratum in the first stage probability, this formula automatically adjusts the probability and weight for any nonresponse and replacements in the stratum. The second stage probability is based on the assumption that one branch is selected in each NSR sample enterprise. In the case of enterprises with only one branch, the second stage probability is equal to 1.

The basic weight for the NSR sample establishments is the inverse of this probability of selection, and can be expressed as follows:

$$W_{Nhi} = \frac{E_{Nh} \times B_{hi}}{n'_h \times E_{hi}},$$

where:

W_{Nhi} = basic weight for the sample branch in the i-th NSR sample enterprise in stratum h

Appendix:

Table 1. STEP Employer Survey Report: Overall Summary of Interview Outcome by Strata

Stratum/Number of firms	Target Sample Size	Reserve Sample	Extra Reserve Sample	Distribution of Firms by Result Code and stratum (for all the visits)								
				1. Completed	Ratio to target sample, %	2.Address is not found	3.The organization doesn't exist	4.The organization refused	5.Ineligible. (on size, or status)	6.The respondent refused	7.The respondent is not available during our survey	8. Other
Republika Srpska	168	291	2	170	101%	36	4	49	3	128	67	4
Rest of Federation	168	265	0	188	112%	53	5	37	4	108	35	3
Sarajevo	168	190	16	178	106%	22	10	25	9	93	33	2
Total	504	746	16	536	106%	111	19	111	16	329	135	9
Actually visited firms					42%	9%	2%	9%	1%	26%	11%	1%

Table 2. Distribution of achieved sample by sector and strata

	Economic activity by sectors	Code	Sarajevo			Other Urban			Share of total, %	Control check: Sample frame %
			Large	Medium	Small	Large	Medium	Small		
A	Agriculture, forestry and fishing	01	1	1	1	3	2	7	2.8	3.0
B	Mining and quarrying	02	0	0	0	1	1	8	1.9	2.4
C	Manufacturing	03	6	11	15	34	39	56	30.0	27.0
D	Electricity, gas, steam and air conditioning supply	04	0	0	2	1	0	5	1.5	3.0
E	Water supply; sewerage, waste management and remediation activities	05	2	0	1	3	9	4	3.5	2.8
F	Construction	06	6	7	7	8	16	6	9.3	8.5
G	Wholesale and retail trade; repair of motor vehicles and motorcycles	07	25	18	15	35	25	16	25.0	26.2
H	Transportation and storage	08	3	3	5	13	8	7	7.3	8.1
I	Accommodation and food service activities	09	4	2	3	4	1	0	2.6	2.0
J	Information and communication	10	4	4	2	6	3	2	3.9	5.0
K	Financial and insurance activities	11	1	0	0	0	0	1	0.4	0.4
L	Real estate activities	12	1	1	1	1	1	0	0.9	1.4
M	Professional, scientific and technical activities	13	8	6	1	12	4	1	6.0	5.0
N	Administrative and support service activities	14	2	4	2	0	2	0	1.9	2.0
O	Public administration and defense; compulsory social security	15	0	0	0	0	0	0	0.0	0.2
P	Education	16	0	1	1	0	1	2	0.9	1.4
Q	Human health and social work activities	17	0	0	1	3	1	2	1.3	1.2
R	Arts, entertainment and recreation	18	0	0	0	1	3	0	0.7	0.4
S	Other service activities	19	0	0	0	0	0	0	0	0.2
T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use	20	0	0	0	0	0	0	0.0	0
	Total	63	58	57	125	116	117	63	100.0%	