Sample Design and Weighting Procedures for Albania STEP Employer Survey

David J. Megill
Sampling Consultant, World Bank
October 2017

1. Sample Design for Albania STEP Employer Survey

The sampling frame for the Albania STEP Employer Survey was based on the business register of all enterprises in Albania with 1 or more employees, which included information on the geographic location, the economic activity, the gender of the owner, and the number of employees. This register is maintained by Institute of Statistics in Albania (INSTAT), who selected the sample firms and replacements for the Albania STEP Survey based on the sampling specifications. The sample design is described in the "STEP Albania Employer Survey Implementation Report". A summary of the sampling plan is described here since the weighting procedures are based on this design.

The firms in the sampling frame were stratified by region, economic activity (based on the NACE code), gender of the owner, and size (in terms of number of employees). The following three regions were defined:

(1) North
(2) Central
(3) South

Four categories were used for the stratification by size:

(1) 1-4 employees
(2) 5-19 employees
(3) 20-99 employees
(4) 100+ employees

In addition to the main sample covering the different economic activities, there was a special interest in particular economic activity groups which were oversampled to ensure sufficient coverage in the survey; this is referred to as the "booster" sample. The following activity groups (with their corresponding activity codes) were identified for this special stratification:

Activities
55 Accommodation
62 Computer programming, consultancy and related activities
79 Travel agency, tour operator reservation service and related activities
1041 Manufacture of oils and fats
1051 Operation of dairies and cheese making
5610 Restaurants and mobile food service activities

The firms in these activities were separated into special strata for the sample allocation and selection.

A stratified two-stage sample design was used for the Albania STEP Employer Survey, with a sample of firms (enterprises) selected at the first stage, and branches selected at the second stage. First it was necessary to allocate the sample by stratum (region, activity, gender and employment size) based on the distribution of the frame and the sample size needed for each domain. The total sample size was 600 firms selected at the first stage, with one sample branch selected in each firm at the second stage. This sample was divided into 400 sample firms for the main strata and 200 sample firms for the oversampled economic activities of special interest (the "booster" sample). This sample size was almost tripled in order to select a reserve of potential replacement firms at the same time. The original sample of 600 firms was selected from this larger first phase sample, and the remaining firms were used as a reserve for selecting replacements for non-responding firms in the same strata. In the case of strata that did not have triple the sample size in the frame, all of the sample enterprises were selected in the first phase. Annex A shows the distribution of the firms and total employment in the full frame as well as the final distribution of the sample of 600 responding firms.

Within each stratum the allocated number of sample firms were selected with probability proportional to size (PPS), where the measure of size for each firm was based on the total number of employees in the frame. There were 4 large firms with a measure of size larger than the sampling interval for the corresponding strata, so these self-representing (SR) firms were selected with a probability of 1 prior to the systematic PPS selection of the remaining non-self-representing (NSR) firms in the stratum. There were also some strata with fewer firms in the frame, where all of the firms were selected during the first phase. In some of these strata there was not a sufficient number of firms for replacement, so it was necessary to select a replacement firm from the next size stratum. In Annex A the distribution of the final sample of firms with completed interviews is based on the actual stratum from which the sample firms were selected.

When each sample firm was contacted, it was determined how many branches (establishments) the firm had. If a sample firm had more than one branch, one branch was selected at random with equal probability to be interviewed. Most of the sample firms only had one branch, but there were many cases of multi-branch sample firms where two-stage sampling was used.

2. Weighting Procedures for Albania STEP Employer Survey

In order for the sample estimates from the Albania STEP Employer Survey data to be representative of the population of firms and branches, it is necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample branch would be equal to the inverse of its overall probability of selection, taking into account each sampling stage.

As described above, a stratified two-stage sample design was used for the Albania STEP Employer Survey. At the first stage a sample of firms was selected in each stratum systematically with PPS, based on the number of employees. At this stage there were 4 large sample firms with a measure of size greater than the initial sampling interval, so they were selected with a probability of 1 because of their size; these are referred to as self-representing (SR) firms. Although it was possible to select more than one branch in a firm with a measure of size that was a multiple of the sampling interval, during the sampling implementation it was only necessary to select one sample branch in each SR firm. Since one branch was selected in each NSR branch, a sample of 600 branches were interviewed in the final sample of 600 firms.

There were also cases of strata with few firms in the frame, for which all the firms were selected at the first phase, including the additional firms for replacement. In this case the original sample firms were selected at the second stage with equal probability.

The weights are specified separately here for the SR and NSR sample firms, and for the strata where all the firms selected at the first phase.

Since one branch was selected in each SR firm, the probability of selection for the sample branch in each SR firm can be expressed as follows:

$$p_{Shi} = \frac{1}{B_{hi}},$$

where:

p_{Shi} = probability of selection for the sample branch in the i-th SR firm in stratum h

B_{hi} = total number of branches identified in the frame for the i-th SR firm 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 firms is the inverse of this probability of selection, and can be expressed as follows:

$$W_{Shi} = B_{hi},$$

where:

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

For each stratum where all of the firms were selected at the first phase, the original sample firms were selected with equal probability, and the replacements that were used were also selected with equal probability. Therefore the probability of selection for the sample firms in these strata can be expressed as follows:

$$p_{Ahi} = \frac{n_{Ah}}{N_{Ah}} \times \frac{1}{B_{hi}},$$

where:

p_{Ahi} = probability of selection for the sample branch in the i-th sample firm in stratum h where all firms in the frame were included in the first phase sample

n_{Ah} = number of completed interviews in stratum h where all firms in the frame were included in first phase sample

N_{Ah} = total number of firms in the frame for stratum h where all firms were included in the first phase sample

B_{hi} = total number of branches identified in the i-th sample firm in stratum h where all firms were included in the first phase sample

This probability includes components from the two stages of selection. In the case of sample firms with only one branch, the second component of this probability would be equal to 1. The weight for these sample firms would be the inverse of this overall probability of selection, and can be expressed as follows:

$$W_{Ahi} = \frac{N_{Ah}}{n_{Ah}} \times B_{hi},$$

where:

W_{Ahi} = weight for the sample branch in the i-th sample firm in stratum h where all firms in the frame were included in first phase sample

The NSR sample firms in each stratum were selected with PPS based on the number of employees in the frame, so the overall 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 firm in stratum h

n_h = number of NSR sample firms selected in stratum h

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

E_{Nh} = total number of employees in the frame for all the NSR firms 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 firm in stratum h

The two components of this probability correspond to the individual sampling stages. In the case of sample firms 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 firm in stratum h

Distribution of firms and total employment in the sampling frame for the Albania STEP Employer Survey by stratum, and distribution of the final sample of completed firm/branch interviews

Stratum	Part of frame	Activity (NACE code)	Gender	Region	Employment size	Number of firms	Total number of employees	Completed sample firm/branch interviews
1	2 Main		1 Male	Central	1 1-4	39,287	60,995	20
2	2 Main		1 Male	Central	2 5-49	5,908	70,876	20
3	2 Main		1 Male	Central	3 50-99	351	24,381	20
4	2 Main		1 Male	Central	4 100+	319	99,413	20
5	2 Main		1 Male	North	1 1-4	9,500	13,734	20
6	2 Main		1 Male	North	2 5-49	1,095	14,288	20
7	2 Main		1 Male	North	3 50-99	46	3,209	20
8	2 Main		1 Male	North	4 100+	39	7,214	20
9	2 Main		1 Male	South	1 1-4	21,881	31,028	20
10	2 Main		1 Male	South	2 5-49	2,037	25,113	20
11	2 Main		1 Male	South	3 50-99	94	6,122	20
12	2 Main		1 Male	South	4 100+	66	17,693	20
13	2 Main		2 Female	Central	1 1-4	20,181	29,851	13
14	2 Main		2 Female	Central	2 5-49	1,625	17,706	13
15	2 Main		2 Female	Central	3 50-99	90	6,550	14
16	2 Main		2 Female	Central	4 100+	82	22,245	14
17	2 Main		2 Female	North	1 1-4	4,165	5,468	19
18	2 Main		2 Female	North	2 5-49	185	2,572	20
19	2 Main		2 Female	North	3 50-99	10	653	8
20	2 Main		2 Female	North	4 100+	8	1,785	7
21	2 Main		2 Female	South	1 1-4	10,331	13,264	16
22	2 Main		2 Female	South	2 5-49	393	4,752	16
23	2 Main		2 Female	South	3 50-99	16	1,124	10
24	2 Main		2 Female	South	4 100+	9	1,608	9

Distribution of firms and total employment in the sampling frame for the Albania STEP Employer Survey by stratum, and distribution of the final sample of completed firm/branch interviews (continued)

Stratum	Part of frame	Activity (NACE code)	Gender	Region	Employment size	Number of firms	Total number of employed	Completed sample firm/branch interviews
25	1 Boost	55	1 Male	Central	1 1-4	192	388	4
26	1 Boost	55	1 Male	Central	2 5-49	118	1,412	3
27	1 Boost	55	1 Male	Central	3 50-99	3	240	2
28	1 Boost	55	1 Male	Central	4 100+	3	496	2
29	1 Boost	55	1 Male	North	1 1-4	289	355	3
30	1 Boost	55	1 Male	North	2 5-49	18	257	3
31	1 Boost	55	1 Male	South	1 1-4	243	374	3
32	1 Boost	55	1 Male	South	2 5-49	63	721	3
33	1 Boost	55	2 Female	Central	1 1-4	79	169	2
34	1 Boost	55	2 Female	Central	2 5-49	40	407	4
35	1 Boost	55	2 Female	Central	3 50-99	1	55	1
36	1 Boost	55	2 Female	Central	4 100+	3	627	2
37	1 Boost	55	2 Female	North	1 1-4	96	121	4
38	1 Boost	55	2 Female	North	2 5-49	8	102	2
39	1 Boost	55	2 Female	South	1 1-4	117	178	3
40	1 Boost	55	2 Female	South	2 5-49	22	201	3
41	1 Boost	55	2 Female	South	3 50-99	1	53	1
42	1 Boost	62	1 Male	Central	1 1-4	484	655	2
43	1 Boost	62	1 Male	Central	2 5-49	66	839	2
44	1 Boost	62	1 Male	Central	3 50-99	5	331	2
45	1 Boost	62	1 Male	Central	4 100+	1	105	
46	1 Boost	62	1 Male	North	1 1-4	52	63	3
47	1 Boost	62	1 Male	North	2 5-49	2	11	1
48	1 Boost	62	1 Male	South	1 1-4	90	116	4
49	1 Boost	62	1 Male	South	2 5-49	2	44	1
50	1 Boost	62	2 Female	Central	1 1-4	120	171	2
51	1 Boost	62	2 Female	Central	2 5-49	11	121	2
52	1 Boost	62	2 Female	North	1 1-4	12	14	3
53	1 Boost	62	2 Female	North	2 5-49	2	25	2
54	1 Boost	62	2 Female	South	1 1-4	21	25	4
55	1 Boost	62	2 Female	South	2 5-49	2	43	
56	1 Boost	79	1 Male	Central	1 1-4	365	598	3
57	1 Boost	79	1 Male	Central	2 5-49	36	466	3
58	1 Boost	79	1 Male	Central	3 50-99	4	248	2

Distribution of firms and total employment in the sampling frame for the Albania STEP Employer Survey by stratum, and distribution of the final sample of completed firm/branch interviews (continued)

Stratum	Part of frame	Activity (NACE code)	Gender	Region	Employment size	Number of firms	Total number of employed	Completed sample firm/branch interviews
59	1 Boost	79	1 Male	North	1 1-4	79	122	2
60	1 Boost	79	1 Male	North	2 5-49	3	25	2
61	1 Boost	79	1 Male	South	1 1-4	91	122	2
62	1 Boost	79	1 Male	South	2 5-49	7	61	2
63	1 Boost	79	2 Female	Central	1 1-4	281	498	3
64	1 Boost	79	2 Female	Central	2 5-49	28	266	3
65	1 Boost	79	2 Female	North	1 1-4	44	71	3
66	1 Boost	79	2 Female	North	2 5-49	1	22	1
67	1 Boost	79	2 Female	South	1 1-4	42	56	3
68	1 Boost	79	2 Female	South	2 5-49	2	19	2
69	1 Boost	1041	1 Male	Central	1 1-4	43	62	4
70	1 Boost	1041	1 Male	Central	2 5-49	3	23	3
71	1 Boost	1041	1 Male	Central	3 50-99	2	132	1
72	1 Boost	1041	1 Male	North	1 1-4	2	2	1
73	1 Boost	1041	1 Male	South	1 1-4	87	131	4
74	1 Boost	1041	1 Male	South	2 5-49	7	53	4
75	1 Boost	1041	2 Female	Central	1 1-4	7	11	4
76	1 Boost	1041	2 Female	South	1 1-4	16	23	5
77	1 Boost	1041	2 Female	South	3 50-99	1	54	1
78	1 Boost	1051	1 Male	Central	1 1-4	36	65	2
79	1 Boost	1051	1 Male	Central	2 5-49	17	216	3
80	1 Boost	1051	1 Male	North	1 1-4	37	68	2
81	1 Boost	1051	1 Male	North	2 5-49	5	56	2
82	1 Boost	1051	1 Male	North	3 50-99	1	59	1
83	1 Boost	1051	1 Male	South	1 1-4	141	224	3
84	1 Boost	1051	1 Male	South	2 5-49	18	234	3
85	1 Boost	1051	1 Male	South	3 50-99	1	52	1
86	1 Boost	1051	1 Male	South	4 100+	2	333	2
87	1 Boost	1051	2 Female	Central	1 1-4	20	36	4
88	1 Boost	1051	2 Female	Central	2 5-49	4	86	2
89	1 Boost	1051	2 Female	North	1 1-4	6	8	3
90	1 Boost	1051	2 Female	North	2 5-49	1	5	1
91	1 Boost	1051	2 Female	South	1 1-4	22	35	3
92	1 Boost	1051	2 Female	South	2 5-49	2	30	2

Distribution of firms and total employment in the sampling frame for the Albania STEP Employer Survey by stratum, and distribution of the final sample of completed firm/branch interviews (continued)

Stratum	Part of frame	Activity (NACE code)	Gender	Region	Employment size	Number of firms	Total number of employed	Completed sample firm/branch interviews
93	1 Boost	5610	1 Male	Central	1 1-4	1,001	1,978	3
94	1 Boost	5610	1 Male	Central	2 5-49	374	3,840	2
95	1 Boost	5610	1 Male	Central	3 50-99	5	363	2
96	1 Boost	5610	1 Male	Central	4 100+	4	762	2
97	1 Boost	5610	1 Male	North	1 1-4	387	702	2
98	1 Boost	5610	1 Male	North	2 5-49	66	588	3
99	1 Boost	5610	1 Male	South	1 1-4	640	1,128	3
100	1 Boost	5610	1 Male	South	2 5-49	128	1,028	3
101	1 Boost	5610	2 Female	Central	1 1-4	389	716	3
102	1 Boost	5610	2 Female	Central	2 5-49	99	916	2
103	1 Boost	5610	2 Female	Central	3 50-99	4	298	2
104	1 Boost	5610	2 Female	North	1 1-4	166	252	3
105	1 Boost	5610	2 Female	North	2 5-49	15	164	2
106	1 Boost	5610	2 Female	South	1 1-4	280	456	2
107	1 Boost	5610	2 Female	South	2 5-49	28	206	2

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

Stratum/ Number of firms	Target Sample Size	Reserve Sample	Extra Reserve Sample	% of original firms completed	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
Central	217	383	217	66%	217	100%	0	29	14	2	10	16	7
North	184	274	184	75%	183	100	6	17	12	1	5	4	3
South	199	333	199	82%	200	100	3	13	12	1	2	4	2
Total	600	990	0	100%	600	100%	9	59	38	4	17	24	12
Actually visited firms					78.6%		1.2%	7.7%	5.0%	0.5%	2.2%	3.1%	1.6%