

The Tajikistan 2013 Enterprise Surveys Data Set

I. Introduction

1. This document provides additional information on the data collected in Tajikistan between February 2013 and April 2014 as part of the fifth round of the Business Environment and Enterprise Performance Survey (BEEPS V), a joint initiative of the World Bank Group (“WB”) and the European Bank for Reconstruction and Development (“EBRD”). It is an enterprise survey whose objective is to gain an understanding of firms’ perception of the environment in which they operate. The survey was until now administered four times at an interval of three years. This has added an important element of dynamics in the study of business environment in transition countries.

The Enterprise Surveys, through interviews with firms in the manufacturing and services sectors, capture business perceptions on the biggest obstacles to enterprise growth, the relative importance of various constraints to increasing employment and productivity, and the effects of a country’s business environment on its international competitiveness. They are used to create statistically significant business environment indicators that are comparable across countries. The Enterprise Surveys are also used to build a panel of enterprise data that will make it possible to track changes in the business environment over time and allow, for example, impact assessments of reforms.

The report outlines and describes the sampling design of the data, the data set structure as well as additional information that may be useful when using the data, such as information on non-response cases and the appropriate use of the weights.

II. Sampling Structure

2. The sample for Tajikistan was selected using stratified random sampling, following the methodology explained in the *Sampling Manual*¹. Stratified random sampling² was preferred over simple random sampling for several reasons³:

a. To obtain unbiased estimates for different subdivisions of the population with some known level of precision.

b. To obtain unbiased estimates for the whole population. The whole population, or universe of the study, is the non-agricultural economy. It comprises: all manufacturing sectors according to the group classification of ISIC Revision 3.1: (group D), construction sector (group F), services sector (groups G and H), and transport, storage, and communications sector (group I). Note that this definition excludes the following sectors: financial intermediation (group J), real estate and renting activities (group K, except sub-sector 72, IT, which was added to the population under study), and all public or utilities-sectors.

c. To make sure that the final total sample includes establishments from all different sectors and that it is not concentrated in one or two of industries/sizes/regions.

1 The complete text can be found at http://www.enterprisesurveys.org/documents/Implementation_note.pdf

2 A stratified random sample is one obtained by separating the population elements into non-overlapping groups, called strata, and then selecting a simple random sample from each stratum. (Richard L. Scheaffer; Mendenhall, W.; Lyman, R., “Elementary Survey Sampling”, Fifth Edition).

3 Cochran, W., 1977, pp. 89; Lohr, Sharon, 1999, pp. 95

d. To exploit the benefits of stratified sampling where population estimates, in most cases, will be more precise than using a simple random sampling method (i.e., lower standard errors, other things being equal.)

e. Stratification may produce a smaller bound on the error of estimation than would be produced by a simple random sample of the same size. This result is particularly true if measurements within strata are homogeneous.

f. The cost per observation in the survey may be reduced by stratification of the population elements into convenient groupings.

3. Three levels of stratification were used in this country: industry, establishment size, and region. The original sample design with specific information of the industries and regions chosen is described in Appendix E.

4. Industry stratification was designed in the way that follows: the universe was stratified into one manufacturing industry, and two service industries (retail, and other services).

5. Size stratification was defined following the standardized definition for the rollout: small (5 to 19 employees), medium (20 to 99 employees), and large (more than 99 employees). For stratification purposes, the number of employees was defined on the basis of reported permanent full-time workers. This seems to be an appropriate definition of the labor force since seasonal/casual/part-time employment is not a common practice, except in the sectors of construction and agriculture.

6. Regional stratification was defined in 4 regions (city and the surrounding business area) throughout Tajikistan.

III. Sampling implementation

7. Given the stratified design, sample frames containing a complete and updated list of establishments as well as information on all stratification variables (number of employees, industry, and region) are required to draw the sample. Great efforts were made to obtain the best source for these listings. However, the quality of the sample frames was not optimal and, therefore, some adjustments were needed to correct for the presence of ineligible units. These adjustments are reflected in the weights computation (*see below*).

8. IPSOS was hired to implement the Tajikistan 2013 enterprise survey. There were local subcontractors in each of the 4 regions surveyed.

9. The sample frame used for the survey in Tajikistan was from: The Statistical Committee of Tajikistan. The database contained the following information

- Coverage;
- Up to datedness;- Availability of detailed stratification variables;
- Contact name(s).

Counts from the sample frame are shown below.

Sample Frame

Source: The Statistical Committee of Tajikistan, 2012

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Dushanbe	5-19	60	122	244	426
	20-99	37	33	101	171
	100+	7	6	32	45
	Total	104	161	377	642
Khatlon	5-19	122	55	174	351
	20-99	58	22	98	178
	100+	23	3	6	32
	Total	203	80	278	561
RRP	5-19	38	35	79	152
	20-99	27	13	62	102
	100+	7	5	10	22
	Total	72	53	151	276
Sughd	5-19	110	55	185	350
	20-99	96	34	124	254
	100+	37	11	17	65
	Total	243	100	326	669
Grand Total		622	394	1132	2148

10. The enumerated establishments were then used as the frame for the selection of a sample with the aim of obtaining interviews at 360 establishments with five or more employees.

11. The quality of the frame was assessed at the onset of the project through visits to a random subset of firms and local contractor knowledge. The sample frame was not immune from the typical problems found in establishment surveys: positive rates of non-eligibility, repetition, non-existent units, etc.

12. Given the impact that non-eligible units included in the sample universe may have on the results, adjustments may be needed when computing the appropriate weights for individual observations. The percentage of confirmed non-eligible units as a proportion of the total number of sampled establishments contacted for the survey was 5.6% (89 out of 1579 establishments)⁴. Breaking down by stratified industries, the following sample targets were achieved (using a4a and a6a):

4 Based on out of target contacts and impossible to contact establishments

Sample design

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Dushanbe	5-19	12	25	33	70
	20-99	9	12	9	30
	100+	5	4	4	13
	Total	26	41	46	113
Khatlon	5-19	17	15	19	51
	20-99	11	8	8	27
	100+	5	2	2	9
	Total	33	25	29	87
RRP	5-19	8	13	5	26
	20-99	6	7	2	15
	100+	2	3	2	7
	Total	16	23	9	48
Sughd	5-19	20	15	20	55
	20-99	15	10	12	37
	100+	10	6	4	20
	Total	45	31	36	112
Grand Total		120	120	120	360

IV. Data Base Structure:

13. The structure of the data base reflects the fact that 3 different versions of the questionnaire were used. The basic questionnaire, the Core Module, includes all common questions asked to all establishments from all sectors. The second expanded variation, the Manufacturing Questionnaire, is built upon the Core Module and adds some specific questions relevant to manufacturing sectors. The third expanded variation, the Retail Questionnaire, is also built upon the Core Module and adds to the core specific questions relevant to retail firms. Each variation of the questionnaire is identified by the index variable, *a0*.

14. All variables are named using, first, the letter of each section and, second, the number of the variable within the section, i.e. *a1* denotes section A, question 1. Variable names preceded by a prefix “ECA” indicate questions specific to the Eastern Europe and Central Asia region, therefore, they may not be found in the implementation of the rollout in other countries. All other suffixed variables are global and are present in all country surveys over the world. All variables are numeric with the exception of those variables with an “x” at the end of their names. The suffix “x” denotes that the variable is alpha-numeric.

15. There are 2 establishment identifiers, *idstd* and *id*. The first is a global unique identifier. The second is a country unique identifier. The variables *a2* (sampling region), *a6a* (sampling establishment’s size), and *a4a* (sampling sector) contain the

establishment's classification into the strata chosen for each country using information from the sample frame. The strata were defined according to the guidelines described above.

16. There are three levels of stratification: industry, size and region. Different combinations of these variables generate the strata cells for each industry/region/size combination. A distinction should be made between the variable *a4a* and *d1a2* (industry expressed as ISIC rev. 3.1 code). The former gives the establishment's classification into one of the chosen industry-strata, whereas the latter gives the actual establishment's industry classification (four digit code) in the sample frame.

17. All of the following variables contain information from the sampling frame. They may not coincide with the reality of individual establishments as sample frames may contain inaccurate information. The variables containing the sample frame information are included in the data set for researchers who may want to further investigate statistical features of the survey and the effect of the survey design on their results.

- a2* is the variable describing sampling regions

- a6a*: coded using the same standard for small, medium, and large establishments as defined above. The code -9 was used to indicate units for which size was undetermined in the sample frame.

- a4a*: coded using ISIC Rev 3.1 codes for the chosen industries for stratification. These codes include most manufacturing industries (15 to 37), retail (52), and (45, 50, 51, 55, 60-64, 72) for other services.

18. The surveys were implemented following a 2 stage procedure. Typically first a screener questionnaire is applied over the phone to determine eligibility and to make appointments. Then a face-to-face interview takes place with the Manager/Owner/Director of each establishment. The variables *a4b* and *a6b* contain the industry and size of the establishment from the screener questionnaire. Variables *a8* to *a11* contain additional information and were also collected in the screening phase.

19. Note that there are additional variables for location (*a3x*) and size (*l1*, *l6* and *l8*) that reflect more accurately the reality of each establishment. Advanced users are advised to use these variables for analytical purposes.

20. Variable *a3x* indicates the actual location of the establishment. There may be divergences between the location in the sampling frame and the actual location, as establishments may be listed in one place but the actual physical location is in another place.

21. Variables *l1*, *l6* and *l8* were designed to obtain a more accurate measure of employment accounting for permanent and temporary employment. Special efforts were made to make sure that this information was not missing for most establishments.

22. Variables *a17x* gives interviewer comments, including problems that occurred during an interview and extraordinary circumstances which could influence results. Please note that sometimes this variable is removed due to privacy issues.

V. Universe Estimates

23. Universe estimates for the number of establishments in each cell in Tajikistan were produced for the strict, median and weak eligibility definitions. The estimates were the multiple of the relative eligible proportions.

24. Appendix B shows the overall estimates of the numbers of establishments in Tajikistan based on the sample frame.

25. For some establishments where contact was not successfully completed during the screening process (because the firm has moved and it is not possible to locate the new location, for example), it is not possible to directly determine eligibility. Thus, different assumptions about the eligibility of establishments result in different adjustments to the universe cells and thus different sampling weights.

26. Three sets of assumptions on establishment eligibility are used to construct sample adjustments using the status code information.

27. Strict assumption: eligible establishments are only those for which it was possible to directly determine eligibility. The resulting weights are included in the variable *wstrict*.

Strict eligibility = (Sum of the firms with codes 1,2,3,4,&16) / Total

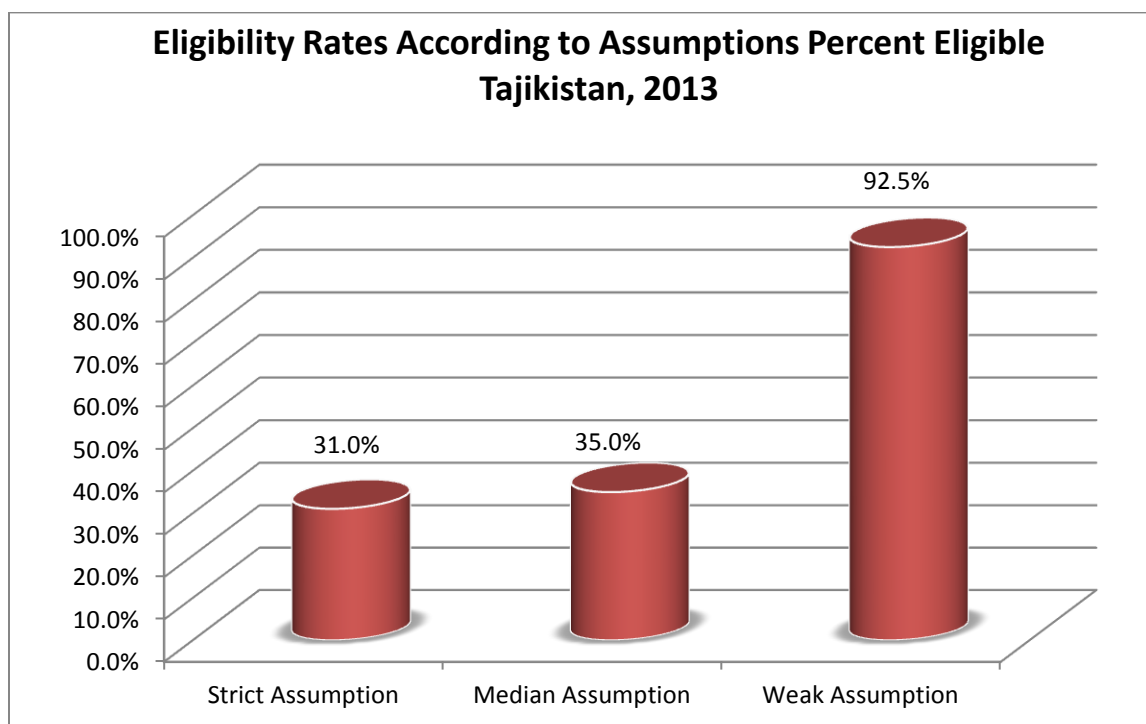
28. Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *wmedian*.

Median eligibility = (Sum of the firms with codes 1,2,3,4,16,10,11, & 13) / Total

29. Weak assumption: in addition to the establishments included in points a and b, all establishments for which it was not possible to contact or that refused the screening questionnaire are assumed eligible. This definition includes as eligible establishments with dead or out of service phone lines, establishments that never answered the phone, and establishments with incorrect addresses for which it was impossible to find a new address. Under the weak assumption only observed non-eligible units are excluded from universe projections. The resulting weights are included in the variable *wweak*.

Weak eligibility = (Sum of the firms with codes 1,2,3,4,16,91,92,93,10,11,12,&13) / Total

30. The indicators computed for the Enterprise Survey website use the median weights. The following graph shows the different eligibility rates calculated for firms in the sample frame under each set of assumptions.



31. Universe estimates for the number of establishments in each industry-region-size cell in Tajikistan were produced for the strict, weak and median eligibility definitions. Appendix D shows the universe estimates of the numbers of registered establishments that fit the criteria of the Enterprise Surveys.

32. Once an accurate estimate of the universe cell projection was made, weights for the probability of selection were computed using the number of completed interviews for each cell.

VI. Weights

33. Since the sampling design was stratified and employed differential sampling, individual observations should be properly weighted when making inferences about the population. Under stratified random sampling, unweighted estimates are biased unless sample sizes are proportional to the size of each stratum. With stratification the probability of selection of each unit is, in general, not the same. Consequently, individual observations must be weighted by the inverse of their probability of selection (probability weights or *pw* in Stata).⁵

34. Special care was given to the correct computation of the weights. It was imperative to accurately adjust the totals within each region/industry/size stratum to account for the presence of ineligible units (the firm discontinued businesses or was unattainable, education or government establishments, establishments with less than 5

⁵ This is equivalent to the weighted average of the estimates for each stratum, with weights equal to the population shares of each stratum.

employees, no reply after having called in different days of the week and in different business hours, no tone on the phone line, answering machine, or fax line⁶, wrong address or moved away and could not get the new references). The information required for the adjustment was collected in the first stage of the implementation: the screening process. Using this information, each stratum cell of the universe was scaled down by the observed proportion of ineligible units within the cell. Once an accurate estimate of the universe cell (projections) was available, weights were computed using the number of completed interviews.

35. Appendix C shows the cell weights for registered establishments in Tajikistan.

VII. Appropriate use of the weights

36. Under stratified random sampling weights should be used when making inferences about the population. Any estimate or indicator that aims at describing some feature of the population should take into account that individual observations may not represent equal shares of the population.

37. However, there is some discussion as to the use of weights in regressions (see Deaton, 1997, pp.67; Lohr, 1999, chapter 11, Cochran, 1953, pp.150). There is not a strong large sample econometric argument in favor of using weighted estimation for a common population coefficient if the underlying model varies per stratum (stratum-specific coefficient): both simple OLS and weighted OLS are inconsistent under regular conditions. However, weighted OLS has the advantage of providing an estimate that is independent of the sample design. This latter point may be quite relevant for the Enterprise Surveys as in most cases the objective is not only to obtain model-unbiased estimates but also design-unbiased estimates (see also Cochran, 1977, pp 200 who favors the use of weighted OLS for a common population coefficient.)⁷

38. From a more general approach, if the regressions are descriptive of the population then weights should be used. The estimated model can be thought of as the relationship that would be expected if the whole population were observed.⁸ If the models are developed as structural relationships or behavioral models that may vary for different parts of the population, then, there is no reason to use weights.

VIII. Non-response

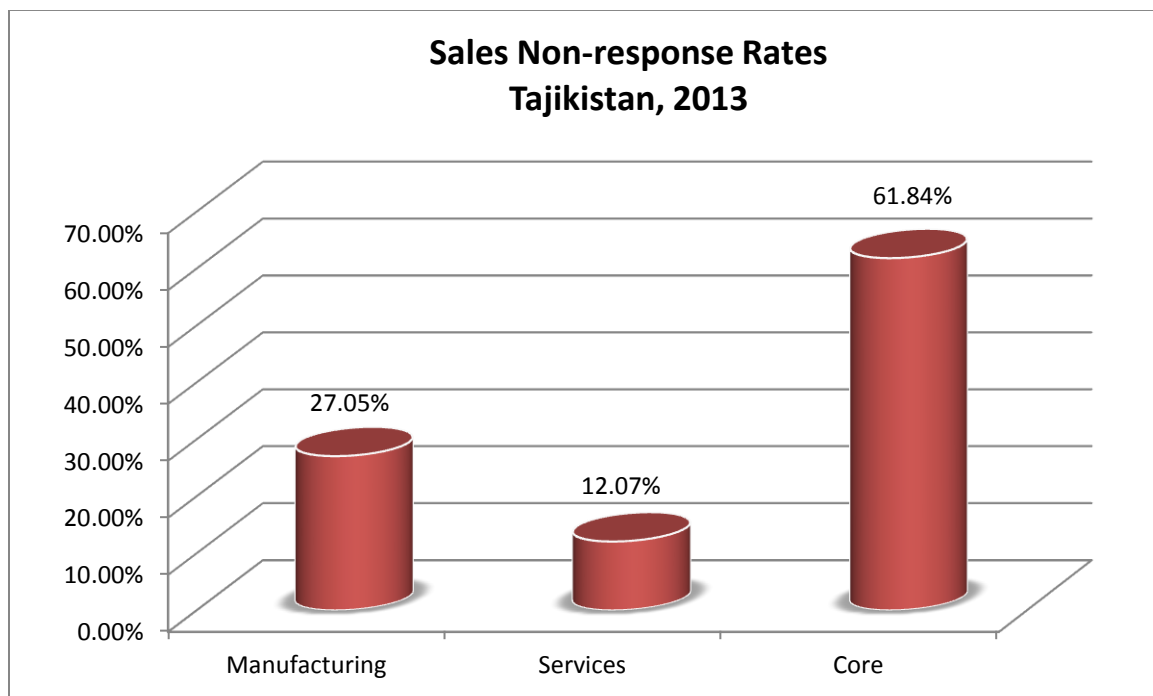
39. Survey non-response must be differentiated from item non-response. The former refers to refusals to participate in the survey altogether whereas the latter refers to the refusals to answer some specific questions. Enterprise Surveys suffer from both problems and different strategies were used to address these issues.

6 For the surveys that implemented a screener over the phone.

7 Note that weighted OLS in Stata using the command `regress` with the option of weights will estimate wrong standard errors. Using the Stata survey specific commands `svy` will provide appropriate standard errors.

8 The use of weights in most model-assisted estimations using survey data is strongly recommended by the statisticians specialized on survey methodology of the JPSM of the University of Michigan and the University of Maryland.

40. Item non-response was addressed by two strategies:
- a- For sensitive questions that may generate negative reactions from the respondent, such as corruption or tax evasion, enumerators were instructed to collect the refusal to respond as a different option from don't know (-8).
 - b- Establishments with incomplete information were re-contacted in order to complete this information, whenever necessary. However, there were clear cases of low response. The following graph shows non-response rates for the sales variable, *d2*, by sector. Please, note that the coding utilized in this dataset does not allow us to differentiate between “Don't know” and “refuse to answer”, thus the non-response in the chart below reflects both categories (DKs and NAs).

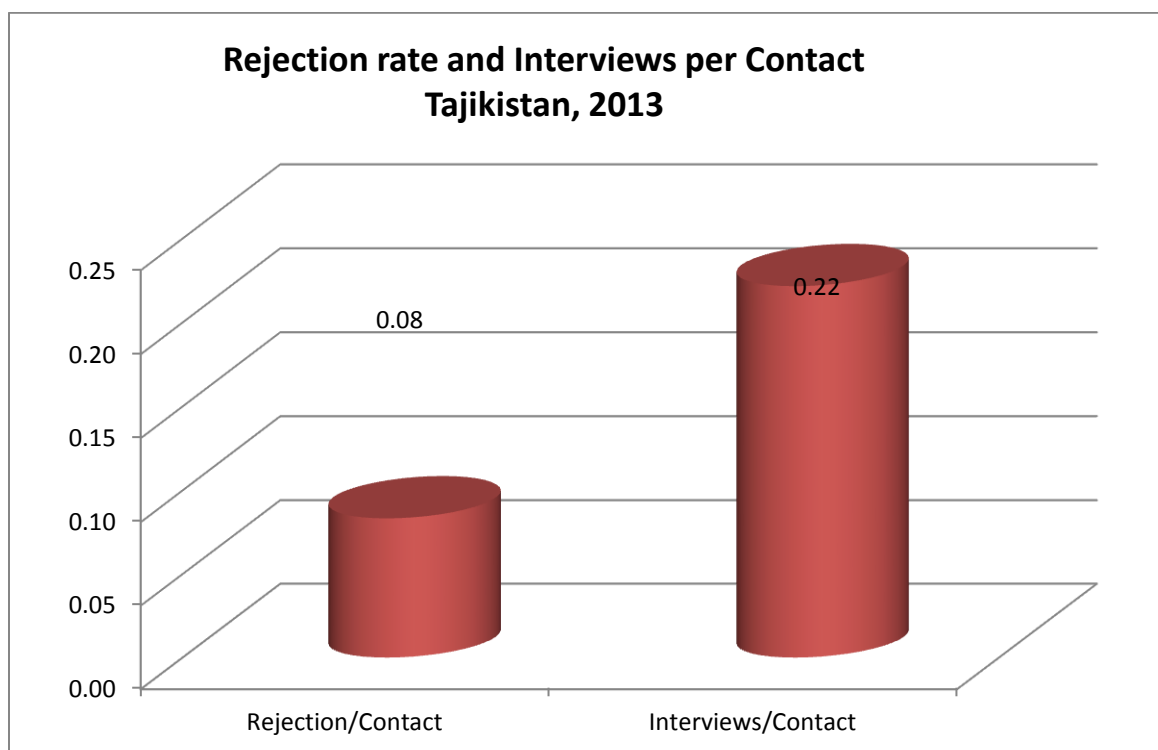


41. Survey non-response was addressed by maximizing efforts to contact establishments that were initially selected for interview. Attempts were made to contact the establishment for interview at different times/days of the week before a replacement establishment (with similar strata characteristics) was suggested for interview. Survey non-response did occur but substitutions were made in order to potentially achieve strata-specific goals. Further research is needed on survey non-response in the Enterprise Surveys regarding potential introduction of bias.

42. As the following graph shows, the number of realized interviews per contacted establishment was 0.22⁹. This number is the result of two factors: explicit refusals to participate in the survey, as reflected by the rate of rejection (which includes rejections of

⁹ The estimate is based on the total number of firms contacted including ineligible establishments.

the screener and the main survey) and the quality of the sample frame, as represented by the presence of ineligible units. The number of rejections per contact was 0.08.



43. Details on the rejection rate, eligibility rate, and item non-response are available at the strata level. This report summarizes these numbers to alert researchers of these issues when using the data and when making inferences. Item non-response, selection bias, and faulty sampling frames are not unique to Tajikistan. All Enterprise Surveys suffer from these shortcomings, but in very few cases they have been made explicit.

References:

Cochran, William G., Sampling Techniques, 1977.

Deaton, Angus, The Analysis of Household Surveys, 1998.

Levy, Paul S. and Stanley Lemeshow, Sampling of Populations: Methods and Applications, 1999.

Lohr, Sharon L. Sampling: Design and Techniques, 1999.

Scheaffer, Richard L.; Mendenhall, W.; Lyman, R., Elementary Survey Sampling, Fifth Edition, 1996.

Appendix A

Status Codes Total:

ELIGIBLES	
1. Eligible establishment (Correct name and address)	494
2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	0
3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	7
4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	2
16. Panel firm - now less than five employees	0
5. The establishment has less than 5 permanent full time employees	5
6. The firm discontinued businesses	36
7. Not a business: private household	0
8. Ineligible activity: education, agriculture, finances, governments...	44
151. Out of target - outside the covered regions, firm moved abroad	0
152. Out of target - firm moved abroad	1
153. Impossible to find	0
91. No reply (<i>after having called in different days of the week and in different business hours</i>)	651
92. Line out of order	53
93. No tone	47
94. Phone number does not exist	81
10. Answering machine	12
11. Fax line - data line	11
12. Wrong address/ moved away and could not get the new references	100
13. Refuses to answer the screener	35
14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	36
Total	1620

Response Outcomes Total:

Complete interviews (<i>Total</i>)	359
Incomplete interviews	0
Eligible in process	50
Refusals	94
Out of target	1
Impossible to contact	955
Ineligible - coop.	1
Refusal to the Screener	41
Total	1620

Status Codes Fresh:

ELIGIBLES	
1. Eligible establishment (<i>Correct name and address</i>)	404
2. Eligible establishment (<i>Different name but same address - the new firm/establishment bought the original firm/establishment</i>)	0
3. Eligible establishment (<i>Different name but same address - the firm/establishment changed its name</i>)	6
4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	2
16. Panel firm - now less than five employees	0
5. The establishment has less than 5 permanent full time employees	3
6. The firm discontinued businesses	28
7. Not a business: private household	0
8. Ineligible activity: education, agriculture, finances, governments...	33
91. No reply (<i>after having called in different days of the week and in different business hours</i>)	563
92. Line out of order	48
93. No tone	40
94. Phone number does not exist	37
10. Answering machine	11
11. Fax line - data line	9
12. Wrong address/ moved away and could not get the new references	78
13. Refuses to answer the screener	35
14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	30
151. Out of target - outside the covered regions, firm moved abroad	0
152. Out of target - firm moved abroad	1
153. Impossible to find	0
Total	1328

Response Outcomes Fresh:

Complete interviews (<i>Total</i>)	292
Incomplete interviews	0
Eligible in process	40
Refusals	80
Out of target	1
Impossible to contact	786
Ineligible - coop.	1
Refusal to the Screener	35
Total	1328

Status Codes Panel:

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	90
	2. Eligible establishment (<i>Different name but same address - the new firm/establishment bought the original firm/establishment</i>)	0
	3. Eligible establishment (<i>Different name but same address - the firm/establishment changed its name</i>)	1
	4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	0
	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	2
	6. The firm discontinued businesses	8
	7. Not a business: private household	0
	8. Ineligible activity: education, agriculture, finances, governments...	11
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	88
	92. Line out of order	5
	93. No tone	7
	94. Phone number does not exist	44
	10. Answering machine	1
	11. Fax line - data line	2
	12. Wrong address/ moved away and could not get the new references	22
	13. Refuses to answer the screener	6
	14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	5
	151. Out of target - outside the covered regions, firm moved abroad	0
	152. Out of target - firm moved abroad	0
	153. Impossible to find	0
Total		292

Response Outcomes Panel:

Complete interviews (<i>Total</i>)	67
Incomplete interviews	0
Eligible in process	10
Refusals	14
Out of target	0
Impossible to contact	169
Ineligible - coop.	0
Refusal to the Screener	6
Total	292

Appendix B

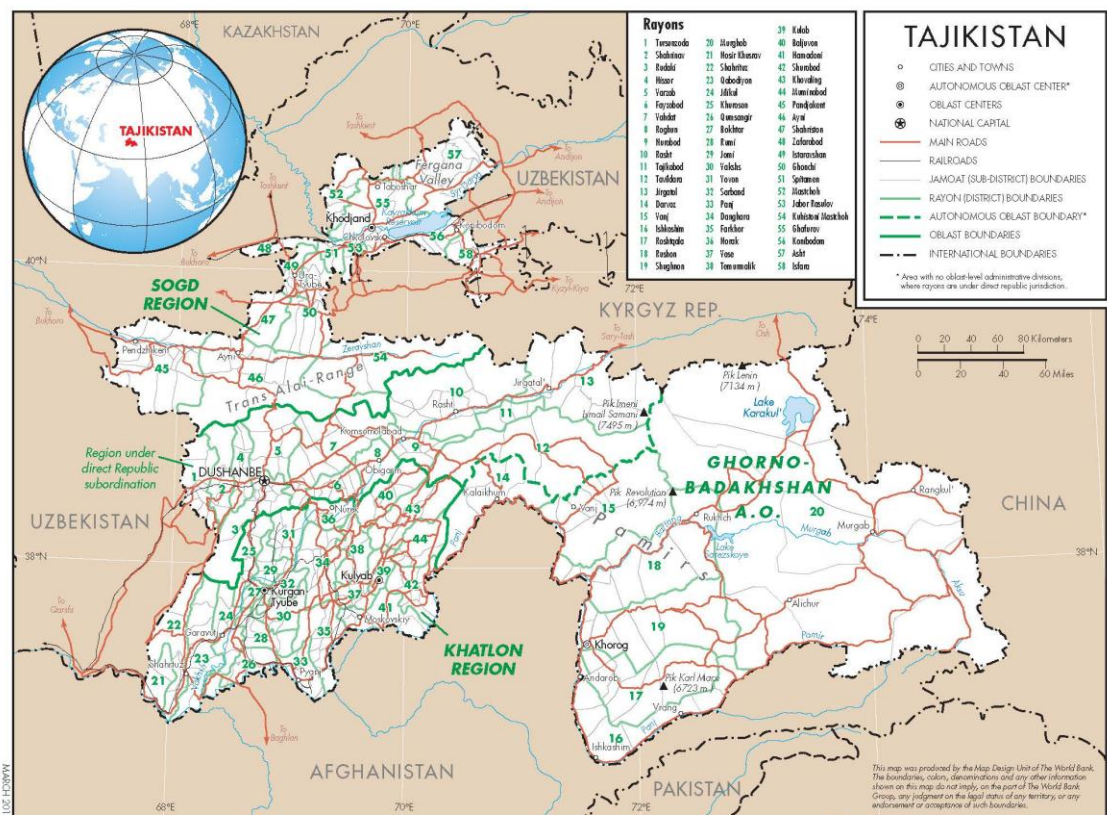
Sampling Frame, Tajikistan:

Source: The Statistical Committee of Tajikistan, 2012

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Dushanbe	5-19	60	122	244	426
	20-99	37	33	101	171
	100+	7	6	32	45
	Total	104	161	377	642
Khatlon	5-19	122	55	174	351
	20-99	58	22	98	178
	100+	23	3	6	32
	Total	203	80	278	561
RRP	5-19	38	35	79	152
	20-99	27	13	62	102
	100+	7	5	10	22
	Total	72	53	151	276
Sughd	5-19	110	55	185	350
	20-99	96	34	124	254
	100+	37	11	17	65
	Total	243	100	326	669
Grand Total		622	394	1132	2148

Appendix C

Tajikistan, administrative divisions



Official provinces	Grouping used for stratification purposes in BEEPS V
Sughd (1)	Sughd (1)
RRP – Region of Republican Subordination (2)	RRP – Region of Republican Subordination (2)
Khatlon (3)	Dushanbe
Gorno-Badakhshan Autonomous Province (4)	Khatlon (3)
	Not covered – only 3% of population

Appendix D

Strict Cell Weights Tajikistan – Panel

Region	Employees	Manufacturing	Retail	Other Services
Dushanbe	5-19	1.0	1.0	
	20-99	1.5		
	100+	1.1		1.0
Khatlon	5-19	1.0		
	20-99			
	100+	1.0		
RRP	5-19			
	20-99			
	100+			
Sughd	5-19	1.0	1.2	1.0
	20-99	1.9		1.0
	100+	2.2		5.0

Strict Cell Weights Tajikistan – Fresh

Region	Employees	Manufacturing	Retail	Other Services
Dushanbe	5-19	4.1	3.9	4.4
	20-99	3.5	1.0	5.7
	100+	8.1	1.0	4.5
Khatlon	5-19	3.1	2.5	4.0
	20-99	5.6	1.8	6.0
	100+	2.3	1.4	1.3
RRP	5-19	4.6	2.0	29.4
	20-99	2.2	2.6	13.3
	100+	5.5	1.0	2.7
Sughd	5-19	3.3	1.8	5.7
	20-99	3.0	1.1	5.3
	100+	2.4	1.0	2.1

Median Cell Weights Tajikistan – Panel

Region	Employees	Manufacturing	Retail	Other Services
Dushanbe	5-19	1.1	1.0	
	20-99	1.7		
	100+	1.2		1.0
Khatlon	5-19	1.0		
	20-99			
	100+	1.1		
RRP	5-19			
	20-99			
	100+			
Sughd	5-19	1.1	1.6	1.0
	20-99	2.0		1.0
	100+	2.2		4.9

Median Cell Weights Tajikistan – Fresh

Region	Employees	Manufacturing	Retail	Other Services
Dushanbe	5-19	4.6	4.5	5.2
	20-99	3.9	1.2	6.8
	100+	9.6	1.0	5.7
Khatlon	5-19	3.1	2.5	4.4
	20-99	5.6	1.9	6.5
	100+	2.5	1.5	1.5
RRP	5-19	5.2	2.4	36.0
	20-99	2.5	3.0	16.3
	100+	6.7	1.0	3.5
Sughd	5-19	3.6	1.9	6.6
	20-99	3.2	1.1	6.1
	100+	2.7	1.0	2.6

Weak Cell Weights Tajikistan - Panel

Region	Employees	Manufacturing	Retail	Other Services
Dushanbe	5-19	3.6		1.9
	20-99	6.1		
	100+	3.0	3.0	
Khatlon	5-19	1.1		
	20-99			
	100+	2.1		
RRP	5-19			
	20-99			
	100+			
Sughd	5-19	2.1	2.1	4.2
	20-99	4.4	2.5	
	100+	3.3	9.3	

Weak Cell Weights Tajikistan – Fresh

Region	Employees	Manufacturing	Retail	Other Services
Dushanbe	5-19	15.5	8.6	16.3
	20-99	14.1	2.4	22.8
	100+	43.5	1.3	24.2
Khatlon	5-19	6.7	3.1	8.8
	20-99	13.1	2.5	14.0
	100+	7.3	2.6	4.2
RRP	5-19	11.6	3.0	74.5
	20-99	5.9	4.0	36.2
	100+	20.1	1.4	9.8
Sughd	5-19	10.9	3.3	18.6
	20-99	10.4	2.2	18.7
	100+	11.1	2.1	10.1

Appendix E

Strict Universe Estimates Tajikistan – Panel

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Dushanbe	5-19	1	5	0	6
	20-99	3	0	0	3
	100+	5	0	1	6
	Total	9	5	1	15
Khatlon	5-19	2	0	0	2
	20-99	0	0	0	0
	100+	2	0	0	2
	Total	4	0	0	4
RRP	5-19	0	0	0	0
	20-99	0	0	0	0
	100+	0	0	0	0
	Total	0	0	0	0
Sughd	5-19	4	2	5	11
	20-99	6	0	6	12
	100+	11	0	5	16
	Total	21	2	16	39
Grand Total		33	7	17	58

Strict Universe Estimates Tajikistan – Fresh

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Dushanbe	5-19	45	79	139	263
	20-99	21	12	51	84
	100+	8	2	13	23
	Total	75	93	204	371
Khatlon	5-19	52	50	85	187
	20-99	28	15	42	84
	100+	7	1	3	11
	Total	87	66	129	282
RRP	5-19	28	31	59	117
	20-99	15	13	27	55
	100+	6	3	5	14
	Total	48	46	91	185
Sughd	5-19	60	32	74	166
	20-99	36	13	37	86
	100+	12	5	6	23
	Total	108	50	118	275
Grand Total		317	255	541	1114

Median Universe Estimates Tajikistan – Panel

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Dushanbe	5-19	1	5	0	6
	20-99	3	0	0	3
	100+	5	0	1	6
	Total	9	5	1	15
Khatlon	5-19	2	0	0	2
	20-99	0	0	0	0
	100+	2	0	0	2
	Total	4	0	0	4
RRP	5-19	0	0	0	0
	20-99	0	0	0	0
	100+	0	0	0	0
	Total	0	0	0	0
Sughd	5-19	4	3	5	13
	20-99	6	0	6	12
	100+	11	0	5	16
	Total	21	3	16	41
Grand Total		35	8	17	60

Median Universe Estimates Tajikistan – Fresh

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Dushanbe	5-19	51	89	167	307
	20-99	23	14	61	99
	100+	10	2	17	29
	Total	84	105	245	435
Khatlon	5-19	53	51	93	197
	20-99	28	15	46	89
	100+	8	2	3	12
	Total	89	68	141	298
RRP	5-19	31	35	72	139
	20-99	17	15	33	65
	100+	7	3	7	17
	Total	55	53	112	220
Sughd	5-19	65	35	85	184
	20-99	38	14	43	95
	100+	14	5	8	26
	Total	116	53	136	306
Grand Total		345	280	634	1259

Weak Universe Estimates Tajikistan – Panel

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Dushanbe	5-19	4	10	0	13
	20-99	12	0	0	12
	100+	12	0	3	15
	Total	28	10	3	40
Khatlon	5-19	2	0	0	2
	20-99	0	0	0	0
	100+	4	0	0	4
	Total	6	0	0	6
RRP	5-19	0	0	0	0
	20-99	0	0	0	0
	100+	0	0	0	0
	Total	0	0	0	0
Sughd	5-19	8	8	10	27
	20-99	13	0	15	28
	100+	17	0	9	26
	Total	38	8	35	81
Grand Total		72	18	38	128

Weak Universe Estimates Tajikistan – Fresh

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Dushanbe	5-19	171	171	523	865
	20-99	85	29	205	319
	100+	44	3	72	119
	Total	299	203	801	1303
Khatlon	5-19	114	63	185	362
	20-99	65	20	98	184
	100+	22	3	8	33
	Total	201	85	292	578
RRP	5-19	70	45	149	264
	20-99	41	20	72	134
	100+	20	4	20	44
	Total	131	69	241	442
Sughd	5-19	196	60	242	498
	20-99	125	26	131	282
	100+	55	10	30	96
	Total	377	96	403	876
Grand Total		1009	454	1736	3199

Appendix F

Original Sample Design, Tajikistan:

Region	Employees	Manufacturing	Retail	Other Services	Grand Total
Dushanbe	5-19	12	25	33	70
	20-99	9	12	9	30
	100+	5	4	4	13
	Total	26	41	46	113
Khatlon	5-19	17	15	19	51
	20-99	11	8	8	27
	100+	5	2	2	9
	Total	33	25	29	87
RRP	5-19	8	13	5	26
	20-99	6	7	2	15
	100+	2	3	2	7
	Total	16	23	9	48
Sughd	5-19	20	15	20	55
	20-99	15	10	12	37
	100+	10	6	4	20
	Total	45	31	36	112
Grand Total		120	120	120	360