

The Pakistan 2013 Enterprise Surveys Data Set

I. Introduction

1. This document provides additional information on the data collected in Pakistan between May 2013 and September 2014. The objective of the Enterprise Survey is to gain an understanding of what firms experience in the private sector.

As part of its strategic goal of building a climate for investment, job creation, and sustainable growth, the World Bank has promoted improving business environments as a key strategy for development, which has led to a systematic effort in collecting enterprise data across countries. The Enterprise Surveys (ES) are an ongoing World Bank project in collecting both objective data based on firms' experiences and enterprises' perception of the environment in which they operate.

The Enterprise Surveys currently cover over 130,000 firms in 135 countries, of which 121 have been surveyed following the standard methodology. This allows for better comparisons across countries and across time. Data 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 Pakistan 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.

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

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.

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 seven manufacturing industries (Food, Textiles, Garments, Chemicals, Non-metallic Minerals, Motor Vehicles, Other Manufacturing) and two service sectors (Retail and Other services).

5. For the Pakistan Enterprise Survey (ES), 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).

6. Regional stratification for the Pakistan ES was defined in five regions: Punjab, Sindh, KPK, Balochistan, and Islamabad

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. Nielsen Pakistan (Lahore) conducted the Pakistan 2013 Enterprise Surveys.

9. The sample frame for Manufacturing establishments was from the Pakistan Bureau of Statistics (PBS). For Retail and Other Services establishments, Nielsen Pakistan provided the sample frame through desk research. For confidentiality purposes, PBS randomly drew the sample of fresh manufacturing establishment to be interviewed based on the sample design provided by the World Bank. The database contained the following information

- Detailed stratification variables;
- Location identifiers- address, phone number, email;
- Contact name(s).

Pakistan ES, Sample Frame

Source: Pakistan Bureau of Statistics (Manufacturing)
Nielsen Pakistan (Lahore) for Retail and Other Services

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Motor Vehicles	Other Manufacturing	Retail	Other Services	Grand Total
Small (5-19)	51	28	37	22	17	22	97	78	219	571
Medium (20-99)	42	53	34	25	12	23	77	15	101	382
Large (100+)	7	61	35	21	13	19	21	14	25	216
Punjab	100	142	106	68	42	64	195	107	345	1169
Small (5-19)	16	5	20	12	3	15	14	50	159	294
Medium (20-99)	56	38	37	17	7	18	26	15	73	287
Large (100+)	15	46	36	23	7	15	27	7	20	196
Sindh	87	89	93	52	17	48	67	72	252	777
Small (5-19)	7	4	3	17	44	2	32	7	37	153
Medium (20-99)	9	3	3	13	7	0	16	7	26	84
Large (100+)	4	3	0	10	3	1	15	2	6	44
KPK	20	10	6	40	54	3	63	16	69	281
Small (5-19)	6	3	0	9	4	3	9	11	21	66
Medium (20-99)	4	8	1	7	3	5	4	4	15	51
Large (100+)	3	7	2	2	2	6	3	2	4	31
Balochistan	13	18	3	18	9	14	16	17	40	148
Small (5-19)	11	0	4	1	17	0	16	21	87	157
Medium (20-99)	19	0	1	3	10	0	14	14	29	90
Large (100+)	4	1	0	1	15	0	10	5	9	45
Islamabad	34	1	5	5	42	0	40	40	125	292
Grand Total	254	260	213	183	164	129	381	252	831	2667

10. The combination of the PBS-provided sample along with the services lists from Nielsen were then used as the sample frame for the Pakistan Enterprise Survey with the aim of obtaining interviews at 1320 establishments.

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. In addition, the sample frame contains no telephone/fax numbers so the local contractor had to screen the contacts by visiting them.

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 4.1% (116 out of 2841 establishments)⁴.

Breaking down by industries and size, the following sample targets were achieved (using screener variables a3a, a4b and a6b):

Achieved sample

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Motor Vehicles	Other Manufacturing	Retail	Other Services	Grand Total
Punjab	85	74	54	32	36	1	131	8	45	466
Micro (<5)	1	1			2		1	1		6
Small (5-19)	56	16	18	11	20		57	4	31	213
Medium (20-99)	24	37	21	10	12	1	49	3	10	167
Large (100+)	4	20	15	11	2		24		4	80
Sindh	31	22	11	8	3	3	23	8	36	145
Micro (<5)	6	1	1				1	3	1	13
Small (5-19)	14	2	1	4	3	1	7	2	4	38
Medium (20-99)	6	3	1	2		1	5	3	13	34
Large (100+)	5	16	8	2		1	10		18	60
Khyber-P	14	7	2	30	53		41	4	27	178
Micro (<5)				1	4		5		9	19
Small (5-19)	9	3	1	11	41		22	4	12	103
Medium (20-99)	2	3	1	13	6		13		6	44
Large (100+)	3	1		5	2		1			12
Balochis	6	8	1	8	2		10		1	36
Small (5-19)	1	2		3	1		3			10
Medium (20-99)	2		1				2			5
Large (100+)	3	6		5	1		5		1	21
Islamaba	23	2	2	3	7		21	16	7	81
Micro (<5)								1		1
Small (5-19)	1		1		2		4	5	1	14
Medium (20-99)	8	1	1	1	3		5	2	2	23
Large (100+)	14	1		2	2		12	8	4	43
Grand Total	159	113	70	81	101	4	226	36	116	906

IV. Data Base Structure:

13. The structure of the data base reflects the fact that 3 different versions of the survey instrument were used for all registered establishments. Questionnaires have common

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

questions (*core* module) and respectfully additional manufacturing and retail specific questions. The eligible manufacturing industries have been surveyed using the **Manufacturing** questionnaire (includes the *core* module, plus manufacturing specific questions). Retail firms have been interviewed using the **Retail** questionnaire (includes the *core* module plus retail specific questions) and the residual eligible services have been covered using the core module only (**Other Services** questionnaire). 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 (some exceptions apply due to comparability reasons). Variable names preceded by a prefix “SAR” indicate questions specific to Pakistan, 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 micro, 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 codes for the chosen industries for stratification. These codes include most manufacturing industries (15 to 37), other manufacturing (2), retail (52), and (45, 50, 51, 55, 60, 63, 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. However, the phone numbers were unavailable in the sample frame, and thus the enumerators applied the screeners in person. 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 variables for 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. 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.

20. 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

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

22. Appendix B shows the overall estimates of the numbers of establishments in Pakistan based on the sample frame.

23. 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.

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

25. 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*.

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

26. 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*.

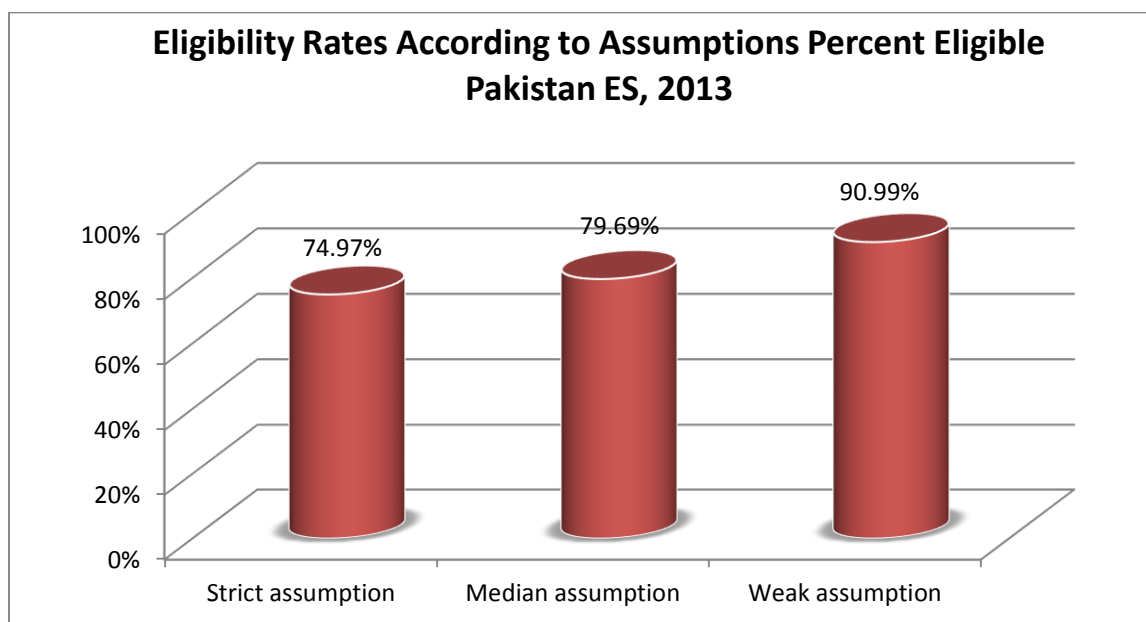
$$\text{Median eligibility} = (\text{Sum of the firms with codes } 1, 2, 3, 4, 16, 10, 11, \& 13) / \text{Total}$$

27. 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 *wwweak*.

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

28. 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.



29. Universe estimates for the number of establishments in each industry-region-size cell in Pakistan 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.

30. 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

31. 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

32. 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 employees, no reply after having called in different days of the week and in different business hours, no tone in the phone line, answering machine, 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.

33. Because sampling size information in the Retail and Other Services sectors turned out to be unreliable, we decided to combine all firm sizes into one overall category (“all sizes > 5”) and sample firms accordingly. Therefore, weights in these two sectors are not size-specific.

34. Appendix C shows the cell weights for registered establishments in Pakistan.

VII. Appropriate use of the weights

35. 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.

36. 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 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 used of weighted OLS for a common population coefficient.)⁷

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

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.

37. 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.

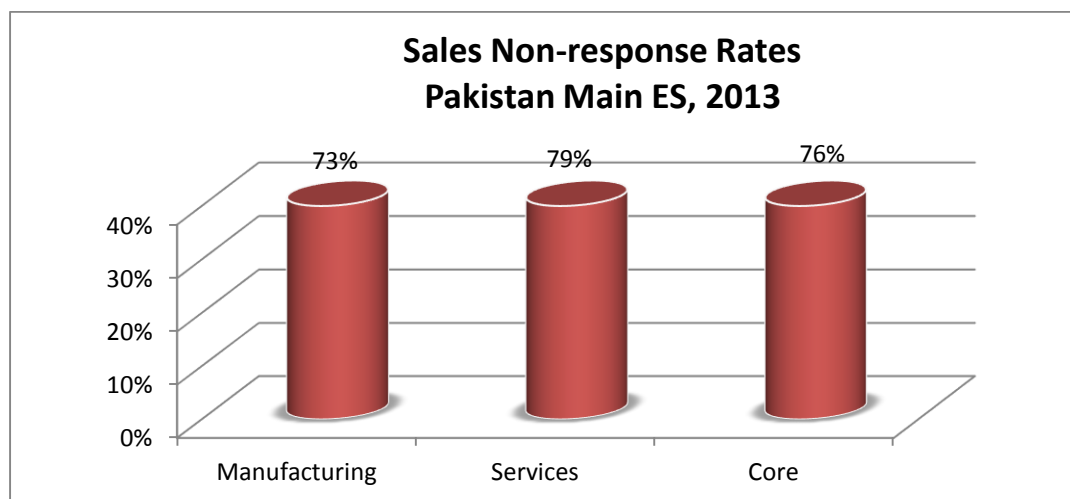
VIII. Non-response

38. 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.

39. 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 (-7).

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 charts below reflect both categories (DKs and NAs).

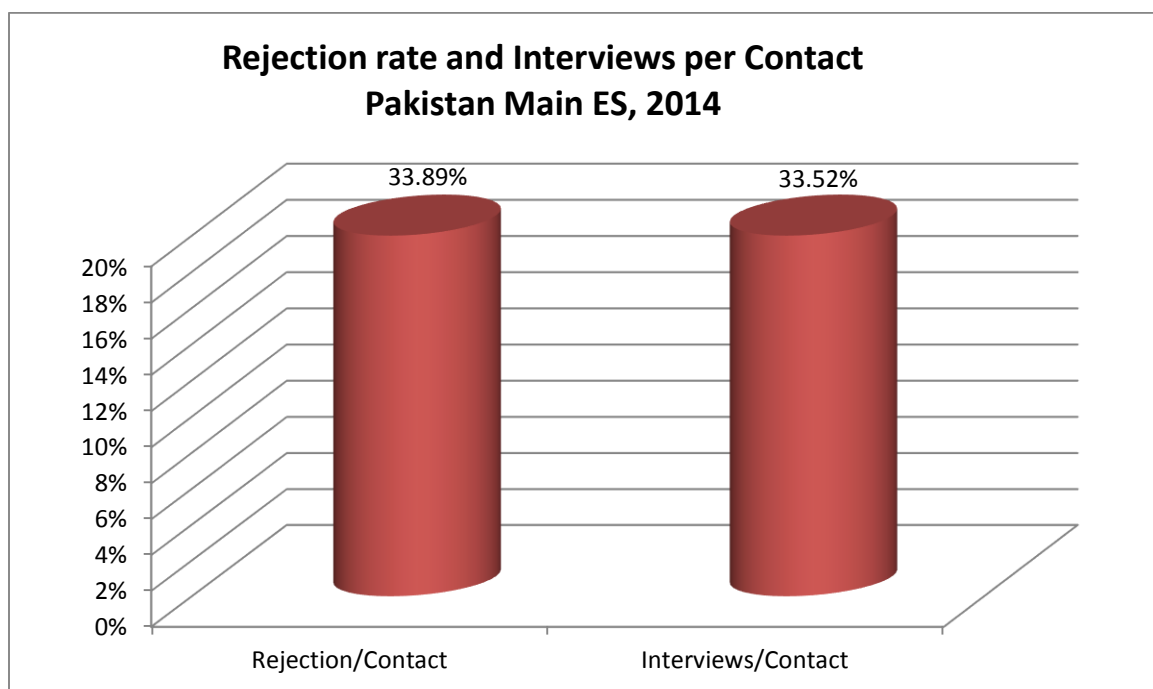


40. Survey non-response was addressed by maximizing efforts to contact establishments that were initially selected for interview. Attempts were made to contact the establishment for interview at different times/days of the week before a replacement establishment (with similar strata characteristics) was suggested for interview. Survey non-

⁸ The use 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.

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.

41. As the following graph shows, the number of interviews per contacted establishments was 0.16⁹. 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 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 0.12.



42. Details on the rejection rate, eligibility rate, and item non-response are available at the level strata. 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 Pakistan. 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.

⁹ The estimate is based on the total no. of firms contacted including ineligible establishments.

Appendix A

Status Codes Enterprise Survey (ES):

	Pakistan
1. Eligible establishment (Correct name and address)	2114
2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	9
3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	2
4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	3
16. Panel firm-now less than five employees	2
5. The establishment has less than 5 permanent full time employees	11
6. The firm discontinued businesses	97
7. Not a business: private household	8
8. Ineligible activity: education, agriculture, finances, governments...	0
91. No reply (<i>after having called in different days of the week and in different business hours</i>)	101
92. Line out of order	0
93. No tone	7
94. Phone number does not exist	50
10. Answering machine	3
11. Fax line - data line	0
12. Wrong address/ moved away and could not get the new references	163
13. Refuses to answer the screener	131
14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	131
151. Out of target - outside the covered regions, firm moved abroad	9
152. Out of target - firm moved abroad	0
153. Out of target - Not registered with statistical agency	0
	2841

Response Outcomes Enterprise (ES) Survey:

	Pakistan
Sample target	1320
Complete interviews (Total)	906
Incomplete interviews	401
Eligible in process	31
Refusals	785
Ineligible	116
Impossible to contact	324
Ineligible - coop.	9
Refusal to the Screener	131
	2703

Response rate	50%
---------------	-----

Out of target + impossible to contact	16%
---------------------------------------	-----

Impossible to contact	12%
-----------------------	-----

Appendix B

Universe Pakistan:

Source: Pakistan Bureau of Statistics (Manufacturing)

Economic Census 2005 of Pakistan (Retail and Other Services). Note these numbers were downward adjusted by the World Bank based on 1) the proportions of services firms to manufacturing firms in each of the 5 provinces (from the Economic Census tabulations) and 2) the discrepancy between the manufacturing totals in the Economic Census 2005 and the PBS provided numbers in December 2012. The 2012 numbers were much lower than the Economic Census 2005 figures.

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Motor Vehicles	Other Manufacturing	Retail	Other Services	Grand Total
Small (5-19)	370	84	60	41	71	30	543	2246	501	3946
Medium (20-99)	318	267	101	47	59	49	455	170	65	1530
Large (100+)	92	303	78	33	23	22	155	63	52	821
Punjab	780	654	239	121	153	101	1153	2478	618	6298
Small (5-19)	174	52	22	11	9	18	60	1054	225	1624
Medium (20-99)	392	176	139	51	25	41	284	246	83	1437
Large (100+)	71	143	77	34	16	30	124	10	3	508
Sindh	637	371	238	96	50	89	468	1310	310	3569
Small (5-19)	165	56	3	39	568	2	215	2818	596	4462
Medium (20-99)	68	30	3	34	127		190	117	47	616
Large (100+)	16	25		15	12	1	37	8	11	125
KPK	249	111	6	88	707	3	442	2943	655	5204
Small (5-19)	94	3		9	19	3	21	731	154	1034
Medium (20-99)	13	24	1	7	4	5	31	72	31	188
Large (100+)	7	13	2	2	2	6	21	0	0	53
Balochistan	114	40	3	18	25	14	73	803	185	1275
Small (5-19)	5			1	17		8	108	33	173
Medium (20-99)	24		1	2	8		23	28	27	113
Large (100+)	3				16		9	5	9	42
Islamabad	32	0	1	3	41	0	40	141	69	328
Grand Total	1812	1176	487	326	976	207	2176	7676	1838	16674

Appendix C

Strict Cell Weights Pakistan - Fresh:

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Motor Vehicles	Other Manufacturing	Retail	Other Services	Retail and Other Services Combined
Punjab										
Small (5-19)	5.99	1.00	1.46	2.39	2.27	10.99	6.43			
Medium (20-99)	7.77	9.15	5.18	2.23	2.66	3.83	9.66			
Large (100+)	15.79	13.50	3.88	4.41	5.49	2.12	15.93			
All Sizes Combined								79.47	7.70	
Sindh										
Small (5-19)	30.79	29.86	3.37	7.35	4.98	13.49	4.39			
Medium (20-99)	60.47	20.95	28.03	9.60		4.22	24.99			
Large (100+)	40.21	19.84		18.81	12.25	4.15	15.36			
All Sizes Combined								51.22	6.99	
KPK										
Small (5-19)	26.13	9.66		2.24	8.36		6.28			
Medium (20-99)	17.07	3.85	1.44	1.67	5.42		20.02			
Large (100+)	2.78			3.06			3.29			
All Sizes Combined										39.52
Balochistan										
Small (5-19)	59.88	1.00		1.30	13.98	1.99				
Medium (20-99)	4.00	5.98		4.17	1.00	1.59				
Large (100+)	1.75	4.30	1.00	1.33	1.00		13.61			
All Sizes Combined										
Islamabad										
Small (5-19)	1.00			1.00	1.01		1.00			
Medium (20-99)	1.14			1.00			4.54			
Large (100+)	1.00				13.20		1.00			
All Sizes Combined								8.69	11.99	

Strict Cell Weights Pakistan - Panel:

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Other Manufacturing	Other Services	Retail and Other Services Combined
Punjab								
Small (5-19)	10.21	14.77	9.36			11.88		
Medium (20-99)	4.97	13.92	7.41	5.94		13.00		
Large (100+)		5.39	14.35			14.53		
All Sizes Combined								
Sindh								
Small (5-19)	23.03		8.88			10.22		
Medium (20-99)	4.86					7.34		
Large (100+)	15.69	5.64		6.74		8.08		
All Sizes Combined							14.20	
KPK								
Small (5-19)								
Medium (20-99)	2.07					5.83		
Large (100+)						1.32		
All Sizes Combined								2.53
Balochistan								
Small (5-19)						7.73		
Medium (20-99)	1.46					2.33		
Large (100+)								
All Sizes Combined							20.65	
Islamabad								
Small (5-19)	2.53		5.33			4.12		
Medium (20-99)	2.08			1.00	1.24	2.07		
Large (100+)						5.51		
All Sizes Combined								

Median Cell Weights Pakistan - Fresh:

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Motor Vehicles	Other Manufacturing	Retail	Other Services	Retail and Other Services Combined
Punjab										
Small (5-19)	6.10	1.00	1.40	2.50	2.39	10.79	6.44			
Medium (20-99)	8.12	9.29	5.12	2.40	2.87	3.86	9.94			
Large (100+)	17.10	14.19	3.98	4.91	6.14	2.22	16.98			
All Sizes Combined								76.61	7.62	
Sindh										
Small (5-19)	33.97	31.98	3.51	8.33	5.68	14.36	4.76			
Medium (20-99)	68.50	23.04	30.01	11.18		4.61	27.86			
Large (100+)	47.17	22.59		22.68	14.85	4.69	17.73			
All Sizes Combined								52.90	7.40	
KPK										
Small (5-19)	29.81	10.70		2.62	9.85		7.05			
Medium (20-99)	20.00	4.38	1.59	2.01	6.56		23.08			
Large (100+)	3.38			3.82			3.93			
All Sizes Combined										111.80
Balochistan										
Small (5-19)	59.08	1.00		1.32	14.26	1.90				
Medium (20-99)	4.06	5.88		4.34	1.00	1.56				
Large (100+)	1.83	4.38	1.00	1.44	1.00		14.06			
All Sizes Combined										
Islamabad										
Small (5-19)	1.00			1.00	1.03		1.00			
Medium (20-99)	1.15			1.00			4.53			
Large (100+)	1.00				14.32		1.00			
All Sizes Combined								8.33	11.80	

Median Cell Weights Pakistan - Panel:

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Other Manufacturing	Other Services	Retail and Other Services Combined
Punjab								
Small (5-19)	11.25	15.05	9.44			12.21		
Medium (20-99)	5.47	14.17	7.46	6.60		13.33		
Large (100+)		5.61	14.78			15.25		
All Sizes Combined								
Sindh								
Small (5-19)	25.52		9.01			10.56		
Medium (20-99)	5.38					7.57		
Large (100+)	17.76	5.91		7.71		8.52		
All Sizes Combined							13.63	
KPK								
Small (5-19)								
Medium (20-99)	3.62					9.51		
Large (100+)						2.20		
All Sizes Combined								3.80
Balochistan								
Small (5-19)						7.58		
Medium (20-99)	1.53					2.28		
Large (100+)								
All Sizes Combined							19.83	
Islamabad								
Small (5-19)	2.66		5.13			4.04		
Medium (20-99)	2.19			1.00	1.16	2.03		
Large (100+)						5.52		
All Sizes Combined								

Weak Cell Weights Pakistan - Fresh:

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Motor Vehicles	Other Manufacturing	Retail	Other Services	Retail and Other Services Combined
Punjab										
Small (5-19)	6.67	1.00	1.55	2.87	2.57	14.22	7.26			
Medium (20-99)	9.03	10.40	5.74	2.79	3.14	5.18	11.39			
Large (100+)	18.78	15.70	4.41	5.65	6.63	2.93	19.23			
All Sizes Combined								121.39	11.74	
Sindh										
Small (5-19)	34.32	32.56	3.58	8.81	5.65	17.48	4.96			
Medium (20-99)	70.37	23.85	31.11	12.03		5.70	29.51			
Large (100+)	47.88	23.10		24.10	14.82	5.74	18.56			
All Sizes Combined								59.57	8.11	
KPK										
Small (5-19)	35.14	12.70		3.24	11.42		8.57			
Medium (20-99)	23.96	5.28	1.93	2.52	7.73		28.51			
Large (100+)	4.00			4.74			4.80			
All Sizes Combined										136.83
Balochistan										
Small (5-19)	76.57	1.00		1.79	18.18	2.96				
Medium (20-99)	5.35	7.81		5.99	1.28	2.48				
Large (100+)	2.39	5.75	1.00	1.96	1.00		18.86			
All Sizes Combined										
Islamabad										
Small (5-19)	1.00			1.00	1.00		1.00			
Medium (20-99)	1.13			1.00			4.58			
Large (100+)	1.00				13.66		1.00			
All Sizes Combined								8.50	11.71	

Weak Cell Weights Pakistan - Panel:

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Other Manufacturing	Other Services	Retail and Other Services Combined
Punjab								
Small (5-19)	12.38	16.04	10.22			13.87		
Medium (20-99)	5.84	14.63	7.83	6.81		14.68		
Large (100+)		5.78	15.45			16.74		
All Sizes Combined								
Sindh								
Small (5-19)	28.48		9.89			12.17		
Medium (20-99)	5.82					8.45		
Large (100+)	19.14	6.17		8.04		9.49		
All Sizes Combined							14.70	
KPK								
Small (5-19)								
Medium (20-99)	5.71					15.52		
Large (100+)						3.57		
All Sizes Combined								5.48
Balochistan								
Small (5-19)						9.31		
Medium (20-99)	1.76					2.72		
Large (100+)								
All Sizes Combined							21.38	
Islamabad								
Small (5-19)	2.72		5.14			4.25		
Medium (20-99)	2.16			1.00	1.05	2.07		
Large (100+)						5.62		
All Sizes Combined								

Appendix D

Strict Universe Estimates - Fresh

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Motor Vehicles	Other Manufacturing	Retail	Other Services	Retail and Other Services Combined	Grand Total
Punjab	591	410	146	73	118	71	668	1431	347	0	3856
Small (5-19)	288	11	23	26	57	22	283	0	0	0	710
Medium (20-99)	241	183	72	25	45	34	290	0	0	0	890
Large (100+)	63	216	51	22	16	15	96	0	0	0	479
All Sizes Combined	0	0	0	0	0	0	0	1431	347	0	1777
Sindh	466	235	183	65	41	64	310	1076	217	0	2656
Small (5-19)	123	30	10	7	5	13	9	0	0	0	198
Medium (20-99)	302	126	112	38	24	30	225	0	0	0	857
Large (100+)	40	79	61	19	12	21	77	0	0	0	309
All Sizes Combined	0	0	0	0	0	0	0	1076	217	0	1292
KPK	147	81	4	56	478	3	291	0	0	869	1928
Small (5-19)	105	39	3	25	384	2	151	0	0	0	708
Medium (20-99)	34	19	1	22	81	0	120	0	0	0	278
Large (100+)	8	23	0	9	12	1	20	0	0	0	73
All Sizes Combined	0	0	0	0	0	0	0	0	0	869	869
Balochistan	71	30	2	12	19	11	42	803	155	0	1145
Small (5-19)	60	3	0	7	14	2	0	0	0	0	85
Medium (20-99)	8	18	1	4	3	3	28	0	0	0	65
Large (100+)	3	9	1	1	2	6	14	0	0	0	36
All Sizes Combined	0	0	0	0	0	0	0	803	155	0	958
Islamabad	11	0	0	3	34	0	24	139	120	0	331
Small (5-19)	2	0	0	1	15	0	6	0	0	0	24
Medium (20-99)	8	0	0	2	6	0	14	0	0	0	30
Large (100+)	1	0	0	0	13	0	4	0	0	0	18
All Sizes Combined	0	0	0	0	0	0	0	139	120	0	259
Grand Total	1287	755	336	208	691	149	1334	3448	838	869	9916

Strict Universe Estimates – Panel

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Other Manufacturing	Other Services	Retail and Other Services Combined	Grand Total
Punjab	45	143	57	23	3	298	35	0	604
Small (5-19)	20	74	28	8	1	178	0	0	310
Medium (20-99)	15	42	15	12	1	91	0	0	175
Large (100+)	10	27	14	3	1	29	0	0	84
All Sizes Combined	0	0	0	0	0	0	35	0	35
Sindh	58	84	37	11	4	80	57	0	331
Small (5-19)	23	18	9	2	3	41	0	0	96
Medium (20-99)	19	27	12	2	1	15	0	0	76
Large (100+)	16	39	16	7	0	24	0	0	102
All Sizes Combined	0	0	0	0	0	0	57	0	57
KPK	24	6	1	2	2	12	0	8	55
Small (5-19)	15	2	0	2	0	4	0	0	23
Medium (20-99)	6	2	1	0	2	6	0	0	17
Large (100+)	3	2	0	0	0	3	0	0	8
All Sizes Combined	0	0	0	0	0	0	0	8	8
Balochistan	18	1	3	1	0	20	21	0	64
Small (5-19)	15	0	2	0	0	15	0	0	32
Medium (20-99)	1	0	0	1	0	2	0	0	5
Large (100+)	2	1	1	0	0	2	0	0	6
All Sizes Combined	0	0	0	0	0	0	21	0	21
Islamabad	28	1	12	3	2	40	0	0	86
Small (5-19)	10	0	11	1	0	21	0	0	42
Medium (20-99)	15	0	1	1	2	8	0	0	27
Large (100+)	3	1	0	1	0	11	0	0	16
All Sizes Combined	0	0	0	0	0	0	0	0	0
Grand Total	174	235	110	40	11	450	112	8	1140

Median Universe Estimates – Fresh:

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Motor Vehicles	Other Manufacturing	Retail	Other Services	Retail and Other Services Combined	Grand Total
Punjab	613	424	146	78	127	72	683	1379	343	0	3865
Small (5-19)	293	11	22	28	60	22	283	0	0	0	718
Medium (20-99)	252	186	72	26	49	35	298	0	0	0	917
Large (100+)	68	227	52	25	18	16	102	0	0	0	507
All Sizes Combined	0	0	0	0	0	0	0	1379	343	0	1722
Sindh	526	261	192	76	45	70	349	1111	230	0	2857
Small (5-19)	136	32	11	8	6	14	10	0	0	0	216
Medium (20-99)	343	138	120	45	24	32	251	0	0	0	953
Large (100+)	47	90	61	23	15	23	89	0	0	0	348
All Sizes Combined	0	0	0	0	0	0	0	1111	230	0	1340
KPK	169	88	5	66	563	3	331	0	0	2460	3685
Small (5-19)	119	43	3	29	453	2	169	0	0	0	818
Medium (20-99)	40	22	2	26	98	0	138	0	0	0	326
Large (100+)	10	23	0	11	12	1	24	0	0	0	81
All Sizes Combined	0	0	0	0	0	0	0	0	0	2460	2460
Balochistan	71	29	2	12	19	11	42	803	155	0	1145
Small (5-19)	59	3	0	7	14	2	0	0	0	0	85
Medium (20-99)	8	18	1	4	3	3	28	0	0	0	65
Large (100+)	4	9	1	1	2	6	14	0	0	0	37
All Sizes Combined	0	0	0	0	0	0	0	803	155	0	958
Islamabad	11	0	0	3	36	0	24	133	118	0	325
Small (5-19)	2	0	0	1	16	0	6	0	0	0	25
Medium (20-99)	8	0	0	2	6	0	14	0	0	0	30
Large (100+)	1	0	0	0	14	0	4	0	0	0	19
All Sizes Combined	0	0	0	0	0	0	0	133	118	0	251
Grand Total	1390	801	344	236	790	156	1429	3426	845	2460	11878

Median Universe Estimates – Panel

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Other Manufacturing	Other Services	Retail and Other Services Combined	Grand Total
Punjab	49	146	58	24	3	307	35	0	622
Small (5-19)	23	75	28	8	1	183	0	0	318
Medium (20-99)	16	43	15	13	1	93	0	0	181
Large (100+)	10	28	15	3	1	31	0	0	87
All Sizes Combined	0	0	0	0	0	0	35	0	35
Sindh	65	86	37	12	4	83	55	0	341
Small (5-19)	26	18	9	2	3	42	0	0	100
Medium (20-99)	22	27	12	2	1	15	0	0	79
Large (100+)	18	41	16	8	0	26	0	0	108
All Sizes Combined	0	0	0	0	0	0	55	0	55
KPK	29	6	1	2	2	18	0	11	69
Small (5-19)	15	2	0	2	0	4	0	0	23
Medium (20-99)	11	2	1	0	2	10	0	0	25
Large (100+)	3	2	0	0	0	4	0	0	9
All Sizes Combined	0	0	0	0	0	0	0	11	11
Balochistan	19	1	3	1	0	19	20	0	63
Small (5-19)	15	0	2	0	0	15	0	0	32
Medium (20-99)	2	0	0	1	0	2	0	0	5
Large (100+)	2	1	1	0	0	2	0	0	6
All Sizes Combined	0	0	0	0	0	0	20	0	20
Islamabad	29	1	11	3	2	39	0	0	86
Small (5-19)	11	0	10	1	0	20	0	0	42
Medium (20-99)	15	0	1	1	2	8	0	0	28
Large (100+)	3	1	0	1	0	11	0	0	16
All Sizes Combined	0	0	0	0	0	0	0	0	0
Grand Total	190	240	110	42	11	467	109	11	1181

Weak Universe Estimates – Fresh:

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Motor Vehicles	Other Manufacturing	Retail	Other Services	Retail and Other Services Combined	Grand Total
Punjab	675	470	162	90	138	96	777	2185	528	0	5121
Small (5-19)	320	11	25	32	64	28	320	0	0	0	800
Medium (20-99)	280	208	80	31	53	47	342	0	0	0	1041
Large (100+)	75	251	57	28	20	21	115	0	0	0	568
All Sizes Combined	0	0	0	0	0	0	0	2185	528	0	2713
Sindh	537	268	196	81	44	86	368	1251	251	0	3083
Small (5-19)	137	33	11	9	6	17	10	0	0	0	222
Medium (20-99)	352	143	124	48	24	40	266	0	0	0	997
Large (100+)	48	92	61	24	15	29	93	0	0	0	362
All Sizes Combined	0	0	0	0	0	0	0	1251	251	0	1502
KPK	200	100	5	83	653	3	406	0	0	3010	4460
Small (5-19)	141	51	3	36	525	2	206	0	0	0	963
Medium (20-99)	48	26	2	33	116	0	171	0	0	0	396
Large (100+)	12	23	0	14	12	1	29	0	0	0	91
All Sizes Combined	0	0	0	0	0	0	0	0	0	3010	3010
Balochistan	92	38	2	17	24	14	47	803	155	0	1192
Small (5-19)	77	3	0	9	18	3	0	0	0	0	110
Medium (20-99)	11	23	1	6	4	5	28	0	0	0	78
Large (100+)	5	12	1	2	2	6	19	0	0	0	46
All Sizes Combined	0	0	0	0	0	0	0	803	155	0	958
Islamabad	11	0	0	3	35	0	24	136	117	0	325
Small (5-19)	2	0	0	1	15	0	6	0	0	0	24
Medium (20-99)	8	0	0	2	6	0	14	0	0	0	30
Large (100+)	1	0	0	0	14	0	4	0	0	0	19
All Sizes Combined	0	0	0	0	0	0	0	136	117	0	253
Grand Total	1516	877	365	274	894	199	1621	4375	1052	3010	14182

Weak Universe Estimates – Panel:

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Other Manufacturing	Other Services	Retail and Other Services Combined	Grand Total
Punjab	52	153	62	25	3	344	35	0	674
Small (5-19)	25	80	31	8	1	208	0	0	353
Medium (20-99)	18	44	16	14	1	103	0	0	194
Large (100+)	10	29	15	3	1	33	0	0	92
All Sizes Combined	0	0	0	0	0	0	35	0	35
Sindh	71	88	38	12	4	94	59	0	366
Small (5-19)	28	18	10	2	3	49	0	0	110
Medium (20-99)	23	27	12	2	1	17	0	0	82
Large (100+)	19	43	16	8	0	28	0	0	115
All Sizes Combined	0	0	0	0	0	0	59	0	59
KPK	35	6	1	2	2	27	0	16	89
Small (5-19)	15	2	0	2	0	4	0	0	23
Medium (20-99)	17	2	1	0	2	16	0	0	38
Large (100+)	3	2	0	0	0	7	0	0	12
All Sizes Combined	0	0	0	0	0	0	0	16	16
Balochistan	19	1	3	1	0	23	21	0	68
Small (5-19)	15	0	2	0	0	19	0	0	36
Medium (20-99)	2	0	0	1	0	3	0	0	5
Large (100+)	2	1	1	0	0	2	0	0	6
All Sizes Combined	0	0	0	0	0	0	21	0	21
Islamabad	29	1	11	3	2	41	0	0	87
Small (5-19)	11	0	10	1	0	21	0	0	43
Medium (20-99)	15	0	1	1	2	8	0	0	27
Large (100+)	3	1	0	1	0	11	0	0	16
All Sizes Combined	0	0	0	0	0	0	0	0	0
Grand Total	206	249	115	43	11	529	115	16	1285

Appendix E

Original Sample Design, Pakistan:

	Food	Textiles	Garments	Chemicals	Non-metalic Minerals	Motor Vehicles	Other Manufacturing	Retail	Other Services	Grand Total
Small (5-19)	37	11	24	19	15	20	41	25	25	217
Medium (20-99)	29	36	28	18	11	21	37	9	21	210
Large (100+)	3	42	25	17	11	17	3	14	7	139
Punjab	69	89	77	54	37	58	81	48	53	566
Small (5-19)	6	3	14	11	3	15	3	18	25	98
Medium (20-99)	36	19	29	15	6	16	18	9	14	162
Large (100+)	3	14	20	13	6	14	3	7	6	86
Sindh	45	36	63	39	15	45	24	34	45	346
Small (5-19)	4	3	3	15	40	2	27	7	5	106
Medium (20-99)	3	3	3	12	6	0	9	7	2	45
Large (100+)	3	3	0	9	3	1	9	2	2	32
KPK	10	9	6	36	49	3	45	16	9	183
Small (5-19)	5	3	0	9	4	3	3	8	2	37
Medium (20-99)	3	7	1	7	3	5	3	4	2	35
Large (100+)	3	6	2	2	2	6	3	2	2	28
Balochistan	11	16	3	18	9	14	9	14	6	100
Small (5-19)	5	0	0	1	17	0	8	10	17	58
Medium (20-99)	7	0	1	2	8	0	7	8	2	35
Large (100+)	3	0	0	0	15	0	6	5	3	32
Islamabad	15	0	1	3	40	0	21	23	22	125
Grand Total	150	150	150	150	150	120	180	135	135	1320

Completed Interviews, Pakistan:

	Food	Textiles	Garments	Chemicals	Non-metallic Minerals	Motor Vehicles	Other Manufacturing	Retail	Other Services	Grand Total
Punjab	85	74	54	32	36	1	131	8	45	466
Micro (<5)	1	1			2		1	1		6
Small (5-19)	56	16	18	11	20		57	4	31	213
Medium (20-99)	24	37	21	10	12	1	49	3	10	167
Large (100+)	4	20	15	11	2		24		4	80
Sindh	31	22	11	8	3	3	23	8	36	145
Micro (<5)	6	1	1				1	3	1	13
Small (5-19)	14	2	1	4	3	1	7	2	4	38
Medium (20-99)	6	3	1	2		1	5	3	13	34
Large (100+)	5	16	8	2		1	10		18	60
Khyber-P	14	7	2	30	53		41	4	27	178
Micro (<5)				1	4		5		9	19
Small (5-19)	9	3	1	11	41		22	4	12	103
Medium (20-99)	2	3	1	13	6		13		6	44
Large (100+)	3	1		5	2		1			12
Balochis	6	8	1	8	2		10		1	36
Small (5-19)	1	2		3	1		3			10
Medium (20-99)	2		1				2			5
Large (100+)	3	6		5	1		5		1	21
Islamaba	23	2	2	3	7		21	16	7	81
Micro (<5)								1		1
Small (5-19)	1		1		2		4	5	1	14
Medium (20-99)	8	1	1	1	3		5	2	2	23
Large (100+)	14	1		2	2		12	8	4	43
Grand Total	159	113	70	81	101	4	226	36	116	906

Appendix F

Challenges and Difficulties in Fieldwork

1. How many enumerators and supervisors worked and when did we start data collection activity
 - a. Enumerators: Overall we started with approximately 40 enumerators. Over the course of the project 83 enumerators worked on this project.
 - b. Supervisors: Total ten (10) supervisors worked on this project.

2. How many trainings provided to them –locations:
We did initial conducted main trainings in the following three stations.
 - a. Lahore
 - b. Karachi
 - c. Rawalpindi/Islamabad
 - d. Hyderabad
 - e. Peshawar
 - f. Multan

Later on follow up trainings were carried out with teams owing to long stretch of fieldwork. The break-up of follow up trainings is as follows:

- a. Lahore (5)
 - b. Karachi (3)
 - c. Rawalpindi/Islamabad (6)
 - d. Hyderabad 2
 - e. Peshawar (3)
 - f. Multan (5)
3. Operational challenges faced & how it had impacted the target to achieve.

The biggest challenge faced in this project was low response rate from the respondents. On an average after at least after every 6th or 7th visit we got a successful interview from majority of the establishments.

Initially finding the establishments was also a big challenge as the addresses were not complete and in certain cases wrong. Therefore it took longer for the teams to find the establishments. Phone numbers were also often wrong therefore team had to go to the location physically to find the required business concern.

Another challenge was that the relevant person was often too busy and it took time to complete the interview.

In Karachi, some of the establishments were very large corporations and getting time from them was a not easy. Further finding the relevant person was even more difficult in such business concerns.

The list was not sufficient according to the quota. After the initial list was exhausted, the second list was received after two months.

Retaining the original team was difficult because the achievement was so low that team got very demoralised.

Benefit of this survey: The respondents often enquired about the benefit they will get out of this and mentioned that the reports we show them are not very useful to them.

4. Information respondents' refused /reluctant to share with Nielsen team.

Balance sheet: Respondents were very reluctant to respond to this question and often refused

Tax related questions were also difficult to administer. This phenomena was especially more prevalent in smaller business concerns

Questions on loans: Respondents were also reluctant to provide answers on loans. Again this was more so in smaller businesses.

Large business establishments: the response rate was very low

5. What are the provinces / regions covered and most challenging geography in terms of implementation of survey

Overall Sindh proved to be most difficult to work in. Hyderabad and Sukkur were two cities that could not be covered completely. Main reason was wrong addresses. The team went to most of the addresses for physical verification and found out that the structure exists but the establishment has moved somewhere.

Quetta: Due to security concern we could not complete 100% of the sample

Karachi: Difficult due to security and corruption from the large establishments