

The Peru 2010 Enterprise Surveys Data Set

I. Introduction

1. This document provides additional information on the data collected in Peru between May 2010 and March 2011 as part of the Latin America and Caribbean (LAC) Enterprise Survey 2010, an initiative of the World Bank.

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 methodology, the sample 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 Peru was selected using stratified random sampling, following the methodology explained in the *Sampling Note*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/locations.

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.)

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

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 location. The original sample design with specific information of the industries and locations chosen is described in Appendix E.

4. Industry stratification was designed in the way that follows: the universe was stratified into 5 manufacturing industries, 1 service industry -retail -, and 1 residual sector as defined in the sampling manual. Three manufacturing industries had targets of 160 interviews; one manufacturing industry (chemicals/rubbers & plastic) had a target of 147. The residual manufacturing, retail, and other services categories each had a target of 120 interviews.

5. Size stratification was defined following the standardized definition for the Enterprise Surveys: 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 four locations (city and the surrounding business area): Lima, Arequipa, Chiclayo and Trujillo.

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 location) 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. TNS Opinion was hired to implement the LAC 2010 enterprise surveys roll out. In Peru the local subcontractor was the Datum International S.A.

9. For Peru, two sample frames were used. The first was supplied by the World Bank and consists of enterprises interviewed in Peru 2006. The World Bank required that attempts should be made to re-interview establishments responding to the Peru 2006 survey where they were within the selected geographical locations and met eligibility criteria. That sample is referred to as the Panel. The second sample frame was obtained from National Institute of Statistics and Informatics (INEI) Economic Census 2007-2008. Each database contained the following information:

- Coverage;
- Up to datedness;
- Availability of detailed stratification variables ;
- Location identifiers- address, phone number, email;
- Electronic format availability;
- Contact name(s).

Counts from sample frames are shown below.

Panel sample counts

Panel Firm Count (Panel sample frame)

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services	Grand Total
Lima	Small	37	79	26	0	1	36	30	209
	Medium	38	42	46	0	1	43	20	190
	Large	26	22	17	0	0	8	24	97
Total		101	143	89	0	2	87	74	496
Arequipa	Small	7	7	3	0	0	14	15	46
	Medium	5	0	2	0	1	11	19	38
	Large	1	2	0	1	0	1	8	13
Total		13	9	5	1	1	26	42	97
Chiclayo	Small	5	2	0	0	0	10	8	25
	Medium	1	0	0	0	0	2	9	12
	Large	0	0	0	0	0	0	0	0
Total		6	2	0	0	0	12	17	37
Trujillo	Small								
	Medium								
	Large								
Total									
Grand Total		120	154	94	1	3	125	133	630

Sample Frames

NOTE: sample drawn by national statistical office; consequently, exact counts for the sample frame are not available.

10. The enumerated establishments were then used as the frame for the selection of a sample with the aim of obtaining interviews at 1000 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. In addition, the sample frame contains no telephone/fax numbers so the local contractor had to screen the contacts by visiting them. Due to response rate and ineligibility issues, additional sample had to be extracted by the World Bank in order to obtain enough eligible contacts and meet the sample targets.

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 5.58% (158 out of 2833 establishments)⁴. Breaking down by stratified industries, the following sample targets were achieved by questionnaire type (using a0 and a6a):

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 (manufacturing, services and IT). The second expanded variation, the Manufacturing Questionnaire, is built upon the Core Module and adds some specific questions relevant to the sector. The third expanded variation, the Services Questionnaire, is also built upon the Core Module and adds to the core specific questions relevant to either retail or IT. 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 “LAC” indicate questions specific to LAC, 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 location), *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 location. Different combinations of these variables generate the strata cells for each industry/location/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

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

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 locations

-*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 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 additional variables for location size by population (*a3*) and firm size by no. of workers (*11*, *16* and *18*) that reflect more accurately the reality of each establishment. Advanced users are advised to use these variables for analytical purposes.

22. Variables *11*, *16* and *18* 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.

23. 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

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

25. Appendix B shows the overall estimates of the numbers of establishments in Peru based on the sample frame.

26. 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.

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

28. Strict assumption: eligible establishments are only those for which it was possible to directly determine eligibility. The resulting weights, which include adjustments applied to panel firms (see below), are included in the variable *w_strict_panadj*.

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

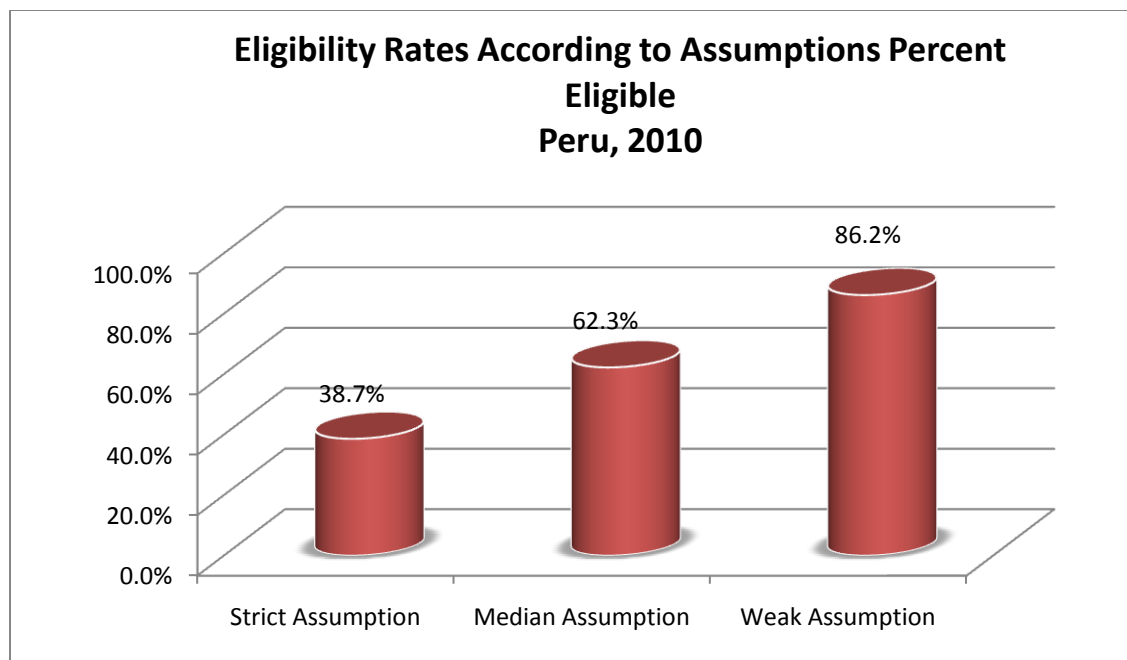
29. Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *w_median_panadj*.

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

30. 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 *w_weak_panadj*.

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

31. 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.



32. Universe estimates for the number of establishments in each industry-location-size cell in Peru 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.

33. 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

34. 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

35. Special care was given to the correct computation of the weights. It was imperative to accurately adjust the totals within each location/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.

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.

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.

The selection of panel firms required additional adjustments to account for varying probabilities of selection between fresh and panel sample universes. For additional information on this methodology, please refer to Enterprise Survey documentation of weighting methodology.

36. Appendix C shows the cell weights for registered establishments in Peru.

VII. Appropriate use of the weights

37. 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.

38. 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

39. 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

40. 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.

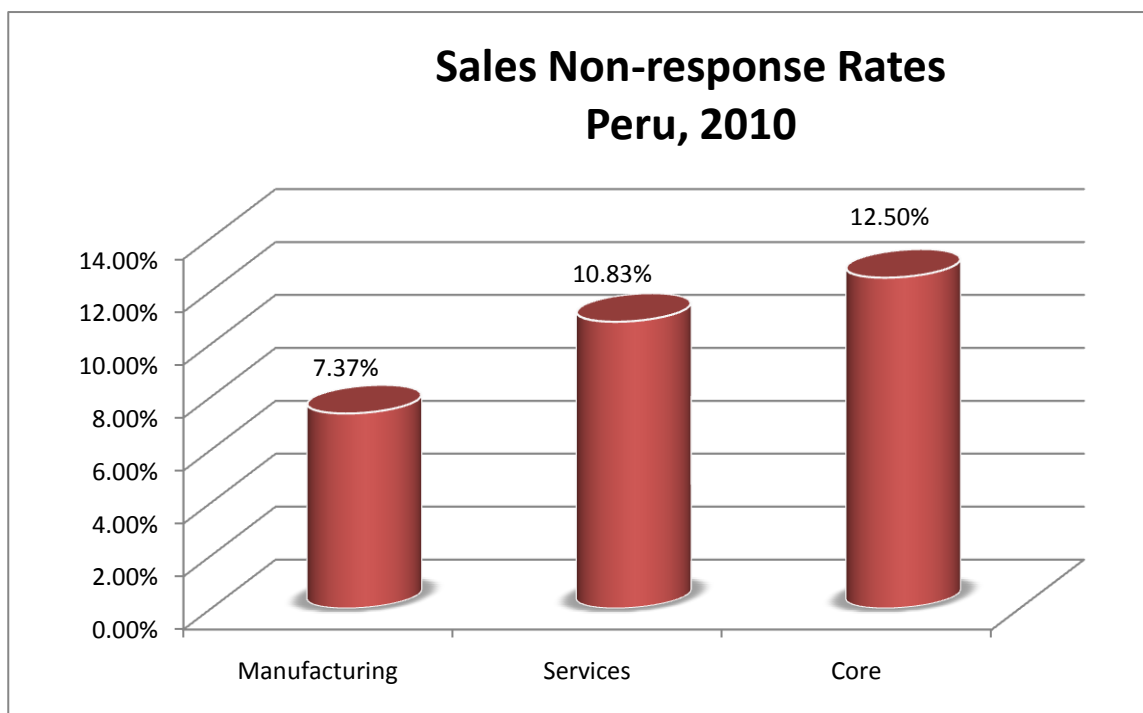
41. Item non-response was addressed by two strategies:

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 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.

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).

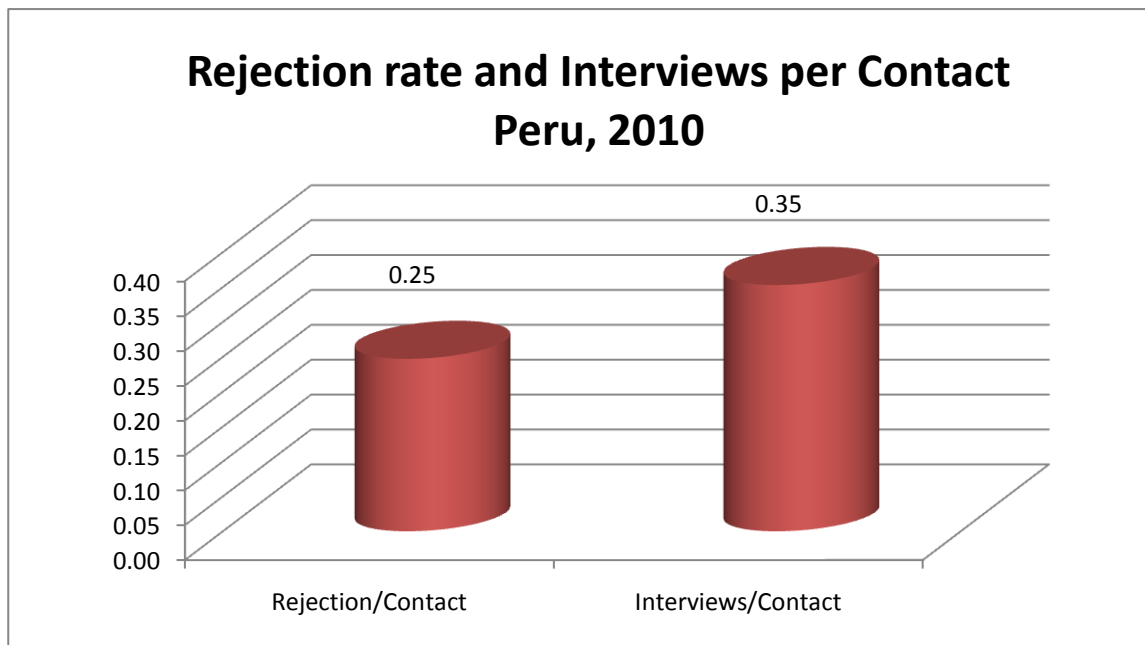
b- Establishments with incomplete information were re-contacted in order to complete this information, whenever necessary. However, there were clear cases of low response. The following graph shows non-response rates for the sales variable, *d2*, by sector. Please, note that the coding utilized in this dataset does not allow us to differentiate between "Don't know" and "refuse to answer", thus the non-response in the chart below reflects both categories (DKs and NAs).



42. 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.

43. As the following graph shows, the number of realized interviews per contacted establishment was 0.35⁹. 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 0.25.

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



44. 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 Peru. 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 Fresh:

	ELIGIBLES	
Eligible	1. Eligible establishment (Correct name and address)	584
	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)	16
	4. Eligible establishment (Moved and traced)	138
Ineligible	5. The establishment has less than 5 permanent full time employees	50
	6 The firm discontinued businesses	61
	7. Not a business: Private household	9
	8. Ineligible activity: Education, Agriculture, Finances, Government, etc.	0
	151 Out of target - outside the covered locations	6
	152. Out of target - moved abroad	0
Unobtainable	91. No reply after having called in different days of the week and in different business hours	85
	92. Line out of order	4
	93. No tone	2
	10. Answering machine	14
	11. Fax line- data line	3
	12. Wrong address/ moved away and could not get the new references	511
	13. Refuses to answer the screener	584
	14. In process (the establishment is being called/ is being contacted - previous to ask the screener)	233
	Total	2300

Response Outcomes Fresh:

Target	
Complete interviews	686
Incomplete interviews	2
Elegible in process	14
Refusals	36
Out of target	126
Impossible to contact	619
Refusal to the Screener	584
TOTAL	2067

Status Codes Panel:

	ELIGIBLES	
Eligible	1. Eligible establishment (Correct name and address)	304
	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)	9
	4. Eligible establishment (Moved and traced)	44
	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	0
	616 The firm discontinued businesses - (Establishment went bankrupt)	18
	618 The firm discontinued businesses - (Original establishment disappeared and is now a different firm)	6
	619 The firm discontinued businesses - (Establishment was bought out by another firm)	3
	620 The firm discontinued businesses - (It was impossible to determine for what reason)	2
	621 The firm discontinued businesses - (Other: SPECIFY in COMMENTS)	0
	7. Not a business: Private household	2
	8. Ineligible activity: Education, Agriculture, Finances, Government, etc.	0
	151 Out of target - outside the covered locations	1
	152. Out of target - moved abroad	0
Unobtainable	91. No reply after having called in different days of the week and in different business hours	56
	92. Line out of order	0
	93. No tone	0
	10. Answering machine	2
	11. Fax line- data line	0
	12. Wrong address/ moved away and could not get the new references	20
	13. Refuses to answer the screener	66
	14. In process (the establishment is being called/ is being contacted - previous to ask the screener)	0
	Total	533

Response Outcomes Panel:

Panel	
Complete interviews	314
Incomplete interviews	1
Elegible in process	28
Refusals	14
Out of target	32
Impossible to contact	78
Refusal to the Screener	66
TOTAL	533

Appendix B

Universe Estimates, Peru:

Location	Size	15	17-18	24-28	27-29	Manuf.	Retail	Services	Total
Arequipa	Small	71	110	6	63	149	1,183	834	2,416
	Medium	12	7	3	9	18	33	73	156
	Large	9	2	4	6	9	6	6	43
Total Arequipa		92	120	13	78	176	1,222	913	2,614
Chiclayo	Small	38	38	4	32	52	360	348	873
	Medium	6	0	1	0	3	17	25	52
	Large	2	0	0	0	0	3	3	8
Total Chiclayo		46	38	6	32	54	380	376	932
Lima	Small	708	1,913	208	521	2,226	8,299	6,228	20,103
	Medium	187	203	127	141	518	411	914	2,501
	Large	111	100	51	29	225	90	190	796
Total Lima		1,006	2,216	386	691	2,968	8,800	7,332	23,399
Trujillo	Small	50	169	15	12	45	403	542	1,235
	Medium	4	6	1	1	6	24	47	90
	Large	11	0	0	1	4	2	5	23
Total Trujillo		65	175	16	14	55	429	594	1,348
Total general		1,209	2,549	421	815	3,253	10,831	9,215	28,293

Appendix C

Strict Cell Weights Peru:

Fresh

Strict Cell Weights*

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services
Lima	Small	26.50	20.52	1.46	3.80	22.73	72.24	161.73
	Medium	3.04	1.74	1.99	1.21	21.81	28.72	115.68
	Large	1.71	1.43	1.58	1.32	3.52	2.16	11.58
Arequipa	Small	1.47	20.52	1.00	1.47	22.73	72.24	161.73
	Medium	1.00	1.00	1.64	1.08	1.58	2.32	6.45
	Large	1.40	1.00	1.00	1.00			4.57
Chiclayo	Small	1.00	20.52	1.00	2.29	22.73	72.24	161.73
	Medium	1.00		1.00		1.21	2.37	1.02
	Large	1.00						1.00
Trujillo	Small	1.00	20.52	1.09	1.00	22.73	72.24	161.73
	Medium	1.00	1.07		1.00	1.35	1.41	1.43
	Large	2.20	1.00		1.00			2.18

*Collapsed cells used in certain cases

Panel

Average Strict Cell Weights, Panel firms

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services
Lima	Small	1.02	1.54	1.00			1.07	1.01
	Medium	1.00	1.52	1.00		1.00	1.26	1.42
	Large	1.11	1.16	1.01			1.06	1.20
Arequipa	Small	1.23	1.48	1.00			1.03	1.00
	Medium	1.09		1.00		1.61	1.00	1.20
	Large	1.00	1.00		1.00		1.00	1.00
Chiclayo	Small	1.00	2.00				1.24	1.00
	Medium	1.00					1.00	1.00
	Large							
Trujillo	Small							
	Medium							
	Large							

Weak Cell Weights Peru:

Fresh

Weak Cell Weights*

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services
Lima	Small	85.61	78.10	4.76	11.51	69.48	334.32	
	Medium	5.97	3.62	3.96	2.24	36.47	72.68	214.97
	Large	3.07	2.72	2.86	2.21	5.37	4.99	19.64
Arequipa	Small	7.07	78.10	1.85	6.60	69.48	334.32	549.64
	Medium	2.78	1.54	4.85	2.96	3.91	8.71	17.79
	Large	3.73	1.00	2.44	2.49			11.52
Chiclayo	Small	3.92	78.10	3.45	9.38	69.48	334.32	549.64
	Medium	2.16		1.81		2.74	8.12	2.55
	Large	1.74						1.34
Trujillo	Small	3.57	78.10	4.48	1.82	69.48	334.32	549.64
	Medium	1.79	2.79		1.91	2.84	4.50	3.33
	Large	4.97	1.00		1.00			4.65

*Collapsed cells used in certain cases

Panel

Average Weak Cell Weights, Panel firms

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services
Lima	Small	2.03	2.83	1.14			2.40	1.90
	Medium	1.37	1.99	1.25		1.00	1.74	1.71
	Large	1.64	1.48	1.87			1.56	1.84
Arequipa	Small	1.88	1.48	1.40			1.13	1.30
	Medium	1.10		1.00		1.91	1.23	1.30
	Large	1.00	1.00		1.41		1.00	1.34
Chiclayo	Small	1.87	2.00				1.44	1.94
	Medium	1.00					1.00	1.25
	Large							
Trujillo	Small							
	Medium							
	Large							

Median Cell Weights Peru:

Fresh

Median Cell Weights*

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services
Lima	Small	50.69	36.26	3.59	7.40	37.95	148.24	
	Medium	5.54	3.02	3.96	2.24	35.82	57.95	190.55
	Large	2.93	2.33	2.86	2.21	5.37	4.09	17.90
Arequipa	Small	2.52	36.26	1.00	2.55	37.95	148.24	270.94
	Medium	1.55	1.00	3.45	1.80	2.31	4.18	9.49
	Large	2.14	1.00	1.79	1.55			6.32
Chiclayo	Small	1.59	36.26	1.78	4.13	37.95	148.24	270.94
	Medium	1.37		1.46		1.84	4.43	1.55
	Large	1.14						1.00
Trujillo	Small	1.52	36.26	2.43	1.00	37.95	148.24	270.94
	Medium	1.20	1.68		1.38	2.01	2.58	2.13
	Large	3.41	1.00		1.00			3.05

*Collapsed cells used in certain cases

Panel

Average Median Cell Weights, Panel firms

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services
Lima	Small	1.12	1.78	1.00			1.50	1.09
	Medium	1.04	1.85	1.00		1.00	1.37	1.47
	Large	1.21	1.21	1.34			1.08	1.42
Arequipa	Small	1.83	1.47	1.14			1.07	1.02
	Medium	1.10		1.00		1.62	1.08	1.21
	Large	1.00	1.00		1.11		1.00	1.09
Chiclayo	Small	1.78	2.00				1.41	1.54
	Medium	1.00					1.00	1.00
	Large							
Trujillo	Small							
	Medium							
	Large							

Appendix D

Strict Universe Estimates

Strict Universe Estimates (Fresh + Panel)

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services	Grand Total
Lima	Small	207.66	99.23	65.62	160.82	276.51	300.74	4.03	1114.60
	Medium	96.82	184.81	74.32	138.14	176.70	168.83	478.35	1317.98
	Large	66.91	73.09	53.59	32.89	144.01	42.07	116.27	528.83
Total		371.39	357.13	193.53	331.85	597.22	511.64	598.64	2961.41
Arequipa	Small	15.48	133.46	5.00	13.20	90.90	225.00	168.73	651.79
	Medium	8.46	4.00	2.00	4.32	9.55	15.28	30.17	73.79
	Large	3.80	3.00	2.00	3.00	1.00	1.00	10.57	24.37
Total		27.75	140.46	9.00	20.53	101.46	241.28	209.48	749.95
Chiclayo	Small	10.00	104.60	23.73	4.58	253.27	511.92	812.67	1720.76
	Medium	3.00	0.00	1.00	0.00	1.21	5.75	12.13	23.09
	Large	2.00	0.00	0.00	0.00	0.00	0.00	2.00	4.00
Total		15.00	104.60	24.73	4.58	254.48	517.66	826.80	1747.84
Trujillo	Small	12.00	225.71	3.28	6.00	111.42	1011.41	1132.14	2501.96
	Medium	2.00	2.14	0.00	0.00	3.70	7.07	17.12	32.03
	Large	4.40	0.00	0.00	1.00	1.00	0.00	3.61	10.01
Total		18.40	227.85	3.28	7.00	116.12	1018.48	1152.87	2544.00
Grand Total		432.54	830.03	230.53	363.96	1069.29	2289.06	2787.80	8003.20

Weak Universe Estimates

Weak Universe Estimates (Fresh + Panel)

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services	Grand Total
Lima	Small	642.89	290.89	180.25	485.68	845.31	1363.70	7.59	3816.32
	Medium	174.42	327.54	131.14	243.53	294.97	398.20	878.66	2448.46
	Large	115.17	128.05	96.50	55.24	220.62	92.68	195.14	903.40
Total		932.48	746.48	407.90	784.45	1360.91	1854.58	1081.39	7168.18
Arequipa	Small	62.25	478.99	8.37	59.39	277.93	1012.04	558.73	2457.70
	Medium	13.85	6.16	2.00	11.85	22.41	42.21	65.10	163.57
	Large	8.46	3.00	4.88	6.38	1.54	1.00	19.57	44.84
Total		84.56	488.15	15.25	77.63	301.88	1055.24	643.40	2666.12
Chiclayo	Small	33.02	392.50	72.94	18.75	777.15	2347.46	2755.95	6397.76
	Medium	5.32	0.00	1.81	0.00	2.74	17.25	25.42	52.53
	Large	5.66	0.00	0.00	0.00	0.00	0.00	2.68	8.34
Total		43.99	392.50	74.74	18.75	779.88	2364.71	2784.06	6458.63
Trujillo	Small	42.79	859.11	13.44	10.95	356.04	4680.53	3847.49	9810.34
	Medium	3.59	5.59	0.00	0.00	7.59	22.49	40.02	79.27
	Large	9.93	0.00	0.00	1.00	1.00	0.00	7.98	19.91
Total		56.31	864.70	13.44	11.95	364.63	4703.02	3895.49	9909.53
Grand Total		1117.33	2491.82	511.33	892.78	2807.30	9977.55	8404.34	26202.46

Median Universe Estimates

Median Universe Estimates (Fresh + Panel)

Location	Firm Size	15	17-18	24-25	27-29	Manuf.	Retail	Services	Grand Total
Lima	Small	382.29	151.43	138.22	313.15	462.83	609.48	4.35	2061.75
	Medium	155.36	294.59	123.75	241.58	289.78	317.09	778.34	2200.50
	Large	104.96	109.14	91.73	55.24	220.62	74.99	175.37	832.05
Total		642.60	555.16	353.70	609.97	973.23	1001.56	958.07	5094.29
Arequipa	Small	25.65	227.85	5.27	22.98	151.81	453.28	278.04	1164.88
	Medium	10.16	4.00	2.00	7.18	14.32	23.21	39.32	100.20
	Large	5.28	3.00	3.57	4.21	1.00	1.00	12.87	30.94
Total		41.09	234.85	10.84	34.37	167.13	477.49	330.24	1296.01
Chiclayo	Small	16.45	183.28	39.74	8.25	423.39	1044.75	1360.84	3076.70
	Medium	3.74	0.00	1.46	0.00	1.84	9.87	16.40	33.30
	Large	2.72	0.00	0.00	0.00	0.00	0.00	2.00	4.72
Total		22.91	183.28	41.20	8.25	425.22	1054.62	1379.24	3114.72
Trujillo	Small	18.22	398.81	7.29	6.00	188.07	2075.37	1896.57	4590.32
	Medium	2.39	3.35	0.00	0.00	5.40	12.89	25.50	49.53
	Large	6.81	0.00	0.00	1.00	1.00	0.00	5.17	13.98
Total		27.42	402.17	7.29	7.00	194.46	2088.26	1927.24	4653.83
Grand Total		734.03	1375.45	413.03	659.59	1760.04	4621.93	4594.78	14158.86

Appendix E

Original Sample Design, Peru:

Emp. Size	Location	15	17-18	24-25	27-29	Other Manuf.	Retail	Other Services	Total
5-19	Lima	20	25	44	53	14	13	9	178
20-99		46	55	51	45	26	30	15	268
100+		46	53	39	35	48	19	20	260
	Regional Total	112	133	134	133	88	62	44	706
5-19	Arequipa	11	10	3	10	9	12	7	62
20-99		7	0	3	2	9	10	10	41
100+		3	4	1	3	3	4	5	23
	Regional Total	21	14	7	15	21	26	22	126
5-19	Chiclayo	5	5	2	3	11	12	13	51
20-99		4		1		2	2	11	20
100+		2					2	3	7
	Regional Total	11	5	3	3	13	16	27	78
5-19	Trujillo	12	7	3	7	6	14	10	59
20-99		2	0	0	1	4	2	12	21
100+		2	1		1	1	0	5	10
	Regional Total	16	8	3	9	11	16	27	90
	TOTAL	160	160	147	160	133	120	120	1000

Completed Interviews, Peru:

Emp. Size	Location	15	17-18	24-25	27-29	Other Manuf.	Retail	Other Services	Total
5-19	Lima	23	27	39	44	14	13	9	169
20-99		40	52	57	53	21	28	16	267
100+		50	56	44	37	43	21	20	271
	Regional Total	113	135	140	134	78	62	45	707
5-19	Arequipa	9	9	3	10	8	12	7	58
20-99		9	2	4	3	6	10	10	44
100+		3	6	1	4	2	4	5	25
	Regional Total	21	17	8	17	16	26	22	127
5-19	Chiclayo	5	5	2	3	11	12	13	51
20-99		4	0	1	0	2	2	11	20
100+		2	0	0	0	0	2	3	7
	Regional Total	11	5	3	3	13	16	27	78
5-19	Trujillo	11	6	3	7	7	12	10	56
20-99		2	0	0	1	4	4	11	22
100+		2	1	0	1	1	0	5	10
	Regional Total	15	7	3	9	12	16	26	88
	TOTAL	160	164	154	163	119	120	120	1000

Appendix F

Local Agency team involved in the study:

Local Agency	DATUM INTERNACIONAL S.A.
Enumerators involved:	16
Other staff involved:	6 tele-operators/coders

Sample Frame:

Characteristic of sample frame used:	Economic Census
Source:	National Institute of Statistics and Informatics (INEI)
Year:	2007/2008
Comments on the quality of sample frame:	Lack of addresses and telephone numbers. This produces a considerable delay in fieldwork.
Year and organism who conducted the last economic census	INEI 2008
Other sources for companies statistics	INEI 2007 Production Ministry 2008

Sectors included in the Sample:

Original Sectors	Manufacturing (Cod. 15, 18, 24, 28, Others), Retail and Core
Added Sectors	Manufacturing (Cod. 15, 17-18, 24-25, 27-29 and Others)

Sample:

Comments on the response rate:	The refusal rate was high. We were not able to contact the total list because many addresses and phone numbers were incorrect or non-existent. Many firms did not want to take part of the survey due to corporate policies. Concerning the Panel list, most of the personnel that participated in the previous wave did not work in the firm anymore.
Other comments:	The sample frame had many repeated companies. The lack of sample caused delays in fieldwork.

Fieldwork:

Date of Fieldwork	Fieldwork start: May 21, 2010 Fieldwork end: March 01, 2011
Problems found during fieldwork:	The attempt was made to contact companies personally (with the aid of field coordinators from the provinces), but in many cases it was unsuccessful as many addresses were wrong or nonexistent. Many companies rejected participating because of security reasons. This

	happened frequently in Trujillo and Chiclayo, due to constant delinquency problems in the regions.
--	--

Questionnaires:

Problems for the understanding of questions (write question number)	<ul style="list-style-type: none"> - A7. The establishment is part of a bigger company. This question was not clear for some respondents. They did not understand if it referred to a group of companies or a company with other establishments. - C15. Over fiscal year [insert last complete fiscal year], did this establishment experience insufficient water supply for production? Some sectors do not depend on water: 17-18, 27-29, Others Manufactures, Retail and Core. - LACL13, LACL15, LACL17. Can you please estimate how many hours were spent on average during the last fiscal year [insert last complete fiscal year]? The interviewers usually calculated based on a “one person” time inversion.
Problems found in the navigability of questionnaires.	Some respondents had trouble answering questions related to finances, or refused to do so. Many respondents refused to give information due to internal corporate policies.
Suggestions or other comments on the questionnaire:	Most of the respondents refuse to give financial data for reasons as company policies.

Country situation

General aspects of economic, political or social situation of the country that could affect the results of the survey:	Constant country growth during previous and current year.
Relevant country events occurred during fieldwork:	Independence Holidays (July 28th and 29th), Municipality General Elections (October 3st), Christmas and New Year.