

# The Bolivia 2010 Enterprise Surveys Data Set

## I. Introduction

1. This document provides additional information on the data collected in Bolivia between June 2010 and October 2010 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 Bolivia was selected using stratified random sampling, following the methodology explained in the *Sampling Note*<sup>1</sup>. Stratified random sampling<sup>2</sup> was preferred over simple random sampling for several reasons<sup>3</sup>:

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.)

---

<sup>1</sup> The complete text can be found at [http://www.enterprisesurveys.org/documents/Implementation\\_note.pdf](http://www.enterprisesurveys.org/documents/Implementation_note.pdf)

<sup>2</sup> 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).

<sup>3</sup> 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 1 manufacturing industry, 1 service industry -retail -, and 1 residual sector as defined in the sampling manual. The manufacturing industry, service industry, and residual sectors had a target each 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 three locations (city and the surrounding business area): La Paz, Santa Cruz, and Cochabamba.

### **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 Bolivia the local subcontractor was Encuestas & Estudios.

9. For Bolivia, two sample frames were used. The first was supplied by the World Bank and consists of enterprises interviewed in Bolivia 2006. The World Bank required that attempts should be made to re-interview establishments responding to the Bolivia 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 produced by Encuestas & Estudios using the Economic Establishments Census, from 2006 updated to 2010. 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 Sample Count (Panel Sample Frame)

Location	Firm Size	Manuf.	Retail	Services	Grand Total
La Paz	Small	122	28	6	156
	Medium	51	22	13	86
	Large	18	6	12	36
Total		191	56	31	278
Santa Cruz	Small	59	30	6	95
	Medium	43	11	12	66
	Large	11	3	15	29
Total		113	44	33	190
Cochabamba	Small	59	19	8	86
	Medium	32	4	5	41
	Large	13	0	5	18
Total		104	23	18	145
Grand Total		408	123	82	613

## Sample Frames

Source: Encuestas & Estudios

Location	Employee size	Manuf.	Retail	Residual	Grand Total
La Paz	Small	1008	496	1262	2766
	Medium	92	64	188	344
	Large	32	0	39	71
La Paz Total		1132	560	1489	3181
Santa Cruz	Small	682	413	888	1983
	Medium	67	60	115	242
	Large	38	2	29	69
Santa cruz Total		787	475	1032	2294
Cochabamba	Small	440	195	465	1100
	Medium	42	32	57	131
	Large	18	1	7	26
Cochabamba Total		500	228	529	1257
Grand Total		2419	1263	3050	6732

10. The two sample frames were then used for the selection of a sample with the aim of obtaining interviews with 360 establishments with five or more employees

11. The quality of the frame was assessed at the outset 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 20.56% (205 out of 997 establishments)<sup>4</sup>.

### 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, retail, and other services). 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

<sup>4</sup> Based on out of target contacts and impossible to contact establishments

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 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 number 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.

20. 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.

21. 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**

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

23. Appendix B shows the overall estimates of the numbers of establishments in Bolivia based on the sample frame.

24. 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.

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

26. 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

27. 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*.

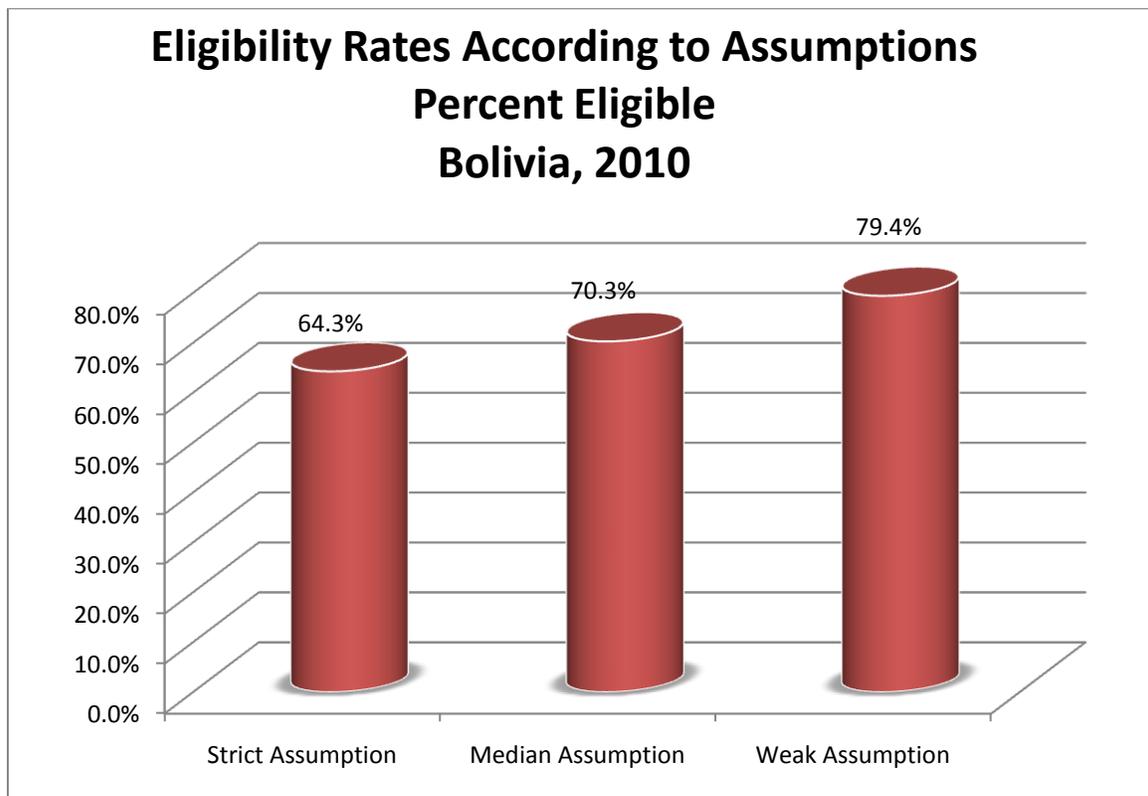
Median eligibility = (Sum of the firms with codes 1,2,3,4,16,10,11, & 13) / Total

28. 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

29. 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.



30. Universe estimates for the number of establishments in each industry-location-size cell in Bolivia 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.

31. 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**

32. 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.)<sup>5</sup>

33. 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 business 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 line<sup>6</sup>, 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.

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.

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

## **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

---

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

<sup>6</sup> For the surveys that implemented a screener over the phone.

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.)<sup>7</sup>

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.<sup>8</sup> 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**

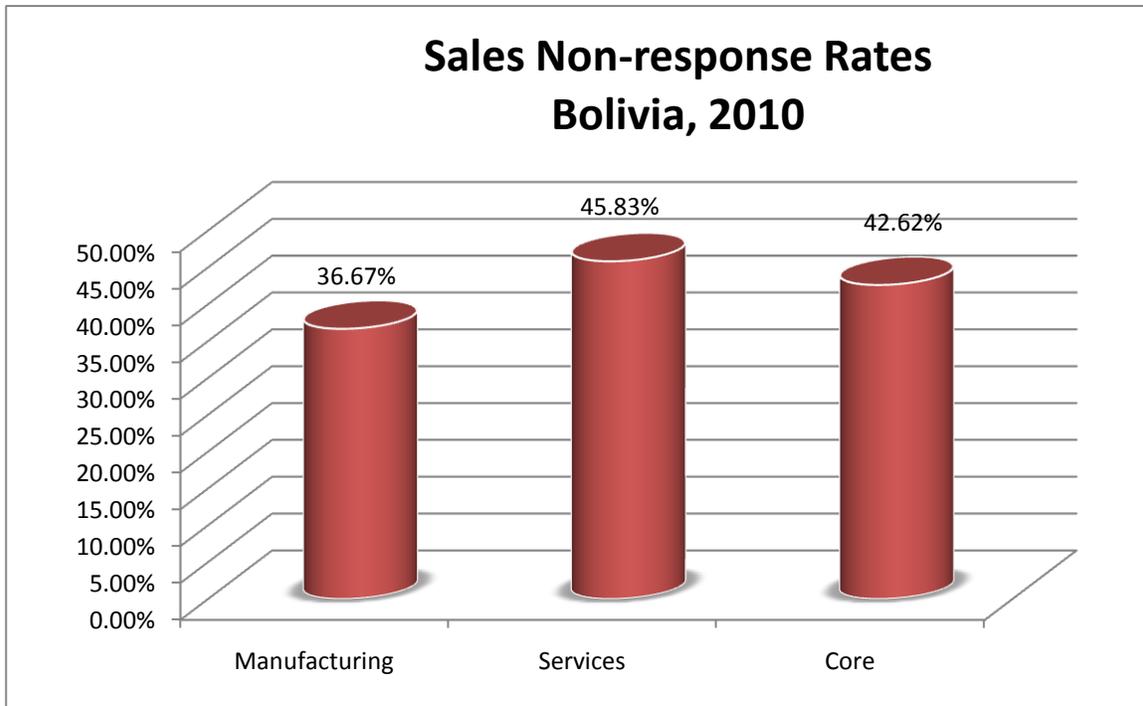
38. Survey non-response must be differentiated from item non-response. The former refers to refusals to participate in the survey altogether whereas the latter refers to the refusals to answer some specific questions. Enterprise Surveys suffer from both problems and different strategies were used to 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 chart below reflects both categories (DKs and NAs).

---

<sup>7</sup> 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.

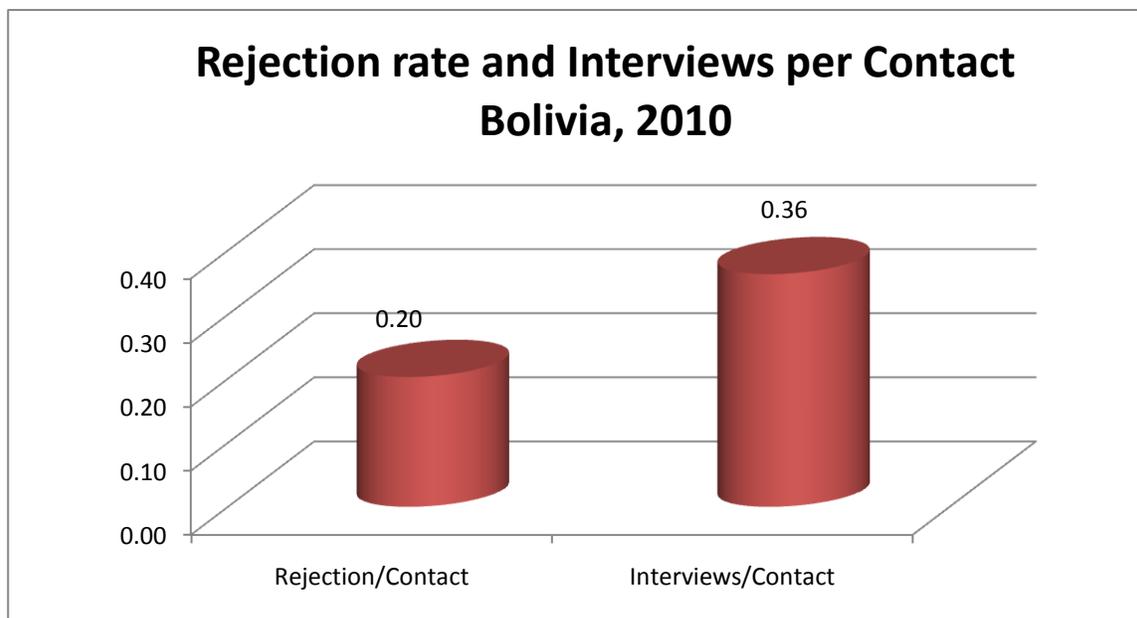
<sup>8</sup> The use weights in most model-assisted estimations using survey data is strongly recommended by the statisticians specialized on survey methodology of the JPSM of the University of Michigan and the University of Maryland.



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-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 realized interviews per contacted establishment was 0.36<sup>9</sup>. 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 0.20.

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



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 Bolivia. 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:

	<b>ELIGIBLES</b>	
Eligible	1. Eligible establishment (Correct name and address)	330
Eligible	2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	2
Eligible	3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	4
Eligible	4. Eligible establishment (Moved and traced)	18
		0
Ineligible	5. The establishment has less than 5 permanent full time employees	14
Ineligible	6 The firm discontinued businesses	7
Ineligible	7. Not a business: Private household	0
Ineligible	8. Ineligible activity: Education, Agriculture, Finances, Government, etc.	146
Ineligible	151 Out of target - outside the covered locations	18
Ineligible	152. Out of target - moved abroad	2
Unobtainable	91. No reply after having called in different days of the week and in different business hours	24
Unobtainable	92. Line out of order	44
Unobtainable	93. No tone	3
Unobtainable	10. Answering machine	8
Unobtainable	11. Fax line- data line	0
Unobtainable	12. Wrong address/ moved away and could not get the new references	30
	13. Refuses to answer the screener	4
	14. In process (the establishment is being called/ is being contacted - previous to ask the screener)	0
	<b>Total</b>	<b>654</b>

### Response Outcomes Fresh:

<b>Target</b>	
Complete interviews	182
Incomplete interviews	3
Elegible in process	41
Refusals	128
Out of target	187
Impossible to contact	109
Refusal to the Screener	4

### Status Codes Panel:

	<b>ELIGIBLES</b>	
Eligible	1. Eligible establishment (Correct name and address)	260
Eligible	2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	1
Eligible	3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	5
Eligible	4. Eligible establishment (Moved and traced)	18
Eligible	16. Panel firm - now less than five employees	3
Ineligible	5. The establishment has less than 5 permanent full time employees	0
Ineligible	616 The firm discontinued businesses - (Establishment went bankrupt)	0
Ineligible	618 The firm discontinued businesses - (Original establishment disappeared and is now a different firm)	0
Ineligible	619 The firm discontinued businesses - (Establishment was bought out by another firm)	1
Ineligible	620 The firm discontinued businesses - (It was impossible to determine for what reason)	2
Ineligible	621 The firm discontinued businesses - (Other: SPECIFY in COMMENTS)	1
Ineligible	7. Not a business: Private household	1
Ineligible	8. Ineligible activity: Education, Agriculture, Finances, Government, etc.	5
Ineligible	151 Out of target - outside the covered locations	8
Ineligible	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	8
Unobtainable	92. Line out of order	12
Unobtainable	93. No tone	0
Unobtainable	10. Answering machine	2
Unobtainable	11. Fax line- data line	0
Unobtainable	12. Wrong address/ moved away and could not get the new references	12
	13. Refuses to answer the screener	4
	14. In process (the establishment is being called/ is being contacted - previous to ask the screener)	0
	<b>Total</b>	<b>343</b>

### Response Outcomes Panel:

<b>Panel</b>	
Complete interviews	180
Incomplete interviews	1
Elegible in process	40
Refusals	66
Out of target	18
Impossible to contact	34
Refusal to the Screener	4

343

## Appendix B

### Universe Estimate, Bolivia:

	Firm Size	<i>Manufacturing</i>	<b>52</b>	<i>Other Services</i>	<b>TOTAL</b>
Cochabamba	5-19 (Small)	528	225	485	1238
	20-99 (Medium)	87	37	67	191
	+100 (Large)	34	1	17	52
	<b>Subtotal</b>	<b>649</b>	<b>263</b>	<b>569</b>	<b>1481</b>
La Paz	5-19 (Small)	1239	561	1276	3076
	20-99 (Medium)	167	92	205	464
	+100 (Large)	67	7	65	139
	<b>Subtotal</b>	<b>1473</b>	<b>660</b>	<b>1546</b>	<b>3679</b>
Santa Cruz	5-19 (Small)	780	451	903	2134
	20-99 (Medium)	118	74	124	316
	+100 (Large)	47	5	39	91
	<b>Subtotal</b>	<b>945</b>	<b>530</b>	<b>1066</b>	<b>2541</b>
<b>Total</b>	<b>Total</b>	<b>3067</b>	<b>1453</b>	<b>3181</b>	<b>7701</b>

## Appendix C

### Strict Cell Weights Bolivia:

#### Panel

Average Strict Cell Weights, Panel firms

Location	Firm Size	Manuf.	Retail	Services
La Paz	Small	1.21	1.48	1.00
	Medium	1.04	1.10	2.51
	Large	1.00	1.14	1.99
Total				
Santa Cruz	Small	1.00	2.13	1.60
	Medium	1.34	1.59	1.57
	Large	1.31	1.00	1.00
Total				
Cochabamba	Small	1.31	1.72	1.00
	Medium	1.65	1.00	1.73
	Large	1.52		1.61

#### Fresh

Strict Cell Weights\*

Location	Firm Size	Manuf.	Retail	Services
La Paz	Small	113.08	29.42	97.98
	Medium	11.78	2.30	3.15
	Large	2.84	2.51	1.41
Total				
Santa Cruz	Small	113.08	26.15	97.98
	Medium	15.27	4.12	7.04
	Large	2.72	2.71	2.71
Total				
Cochabamba	Small	113.08	43.95	97.98
	Medium	10.04	3.27	31.41
	Large	2.99		1.65

\*Collapsed cells used in certain cases

## Weak Cell Weights Bolivia:

### Panel

Average Weak Cell Weights, Panel firms

Location	Firm Size	Manuf.	Retail	Services
La Paz	Small	1.28	2.38	1.00
	Medium	1.16	1.30	2.99
	Large	1.00	1.25	2.61
Total				
Santa Cruz	Small	1.09	2.41	2.33
	Medium	2.22	1.58	1.85
	Large	2.16	1.00	1.00
Total				
Cochabamba	Small	1.68	2.19	1.18
	Medium	2.24	1.33	2.11
	Large	2.15		1.54

### Fresh

Weak Cell Weights\*

Location	Firm Size	Manuf.	Retail	Services
La Paz	Small	128.23	38.21	111.22
	Medium	13.53	2.85	3.62
	Large	3.18	3.04	1.58
Total				
Santa Cruz	Small	128.23	29.99	111.22
	Medium	15.48	4.51	7.14
	Large	2.72	2.90	2.71
Total				
Cochabamba	Small	128.23	51.50	111.22
	Medium	10.80	3.67	34.27
	Large	3.11		1.72

\*Collapsed cells used in certain cases

## Median Cell Weights Bolivia:

### Panel

Average Median Cell Weights, Panel firms

Location	Firm Size	Manuf.	Retail	Services
La Paz	Small	1.13	1.61	1.00
	Medium	1.05	1.11	2.55
	Large	1.02	1.07	1.98
Total				
Santa Cruz	Small	1.00	2.12	1.61
	Medium	1.39	1.59	1.85
	Large	1.30	1.00	1.00
Total				
Cochabamba	Small	1.29	1.95	1.00
	Medium	1.87	1.00	1.90
	Large	2.15		1.54

### Fresh

Median Cell Weights\*

Location	Firm Size	Manuf.	Retail	Services
La Paz	Small	115.76	30.15	99.62
	Medium	12.04	2.34	3.20
	Large	3.06	2.70	1.51
Total				
Santa Cruz	Small	115.76	27.02	99.62
	Medium	15.48	4.23	7.14
	Large	2.72	2.90	2.71
Total				
Cochabamba	Small	115.76	43.95	99.62
	Medium	10.04	3.27	31.41
	Large	3.10		1.70

\*Collapsed cells used in certain cases

## Appendix D

### Strict Universe Estimates

Strict Universe Estimates (Fresh + Panel)

Location	Firm Size	Manuf.	Retail	Services	Grand Total
La Paz	Small	691.84	310.46	393.93	1396.23
	Medium	84.23	60.09	85.61	229.94
	Large	35.54	7.08	24.83	67.45
Total		811.61	377.63	504.37	1693.62
Santa Cruz	Small	684.50	315.32	226.91	1226.73
	Medium	74.46	60.58	60.28	195.32
	Large	37.93	5.71	25.54	69.18
Total		796.89	381.61	312.73	1491.22
Cochabamba	Small	230.09	145.61	395.93	771.63
	Medium	50.05	73.15	38.33	161.53
	Large	35.58	0.00	11.44	47.02
Total		315.71	218.76	445.70	980.18
Grand Total		1924.21	978.01	1262.80	4165.02

### Weak Universe Estimates

Weak Universe Estimates (Fresh + Panel)

Location	Firm Size	Manuf.	Retail	Services	Grand Total
La Paz	Small	783.44	408.33	446.88	1638.65
	Medium	96.23	73.39	99.28	268.90
	Large	38.67	8.04	29.41	76.12
Total		918.33	489.77	575.57	1983.67
Santa Cruz	Small	775.93	361.19	259.41	1396.52
	Medium	84.12	65.25	62.94	212.31
	Large	41.36	5.90	25.54	72.79
Total		901.40	432.34	347.88	1681.62
Cochabamba	Small	261.50	172.01	449.60	883.12
	Medium	55.82	84.83	42.69	183.35
	Large	40.18	0.00	11.48	51.67
Total		357.51	256.85	503.78	1118.14
Grand Total		2177.24	1178.96	1427.23	4783.43

## Median Universe Estimates

Median Universe Estimates (Fresh + Panel)

Location	Firm Size	Manuf.	Retail	Services	Grand Total
La Paz	Small	707.04	319.26	400.50	1426.79
	Medium	85.89	60.90	86.82	233.60
	Large	37.67	6.97	25.99	70.63
Total		830.60	387.12	513.31	1731.03
Santa Cruz	Small	700.59	324.83	231.10	1256.51
	Medium	75.79	61.89	62.94	200.61
	Large	37.89	5.90	25.54	69.33
Total		814.27	392.62	319.57	1526.45
Cochabamba	Small	235.40	147.44	402.50	785.34
	Medium	50.90	73.15	39.01	163.06
	Large	39.38	0.00	11.41	50.79
Total		325.68	220.59	452.92	999.19
Grand Total		1970.54	1000.33	1285.80	4256.67

## Appendix E

### Original Sample Design, Bolivia:

Emp. Size	Location	Manufacturing	Retail	Other Services	
5-19	La Paz	15	26	15	56
20-99		14	26	16	56
100+		23	2	31	56
	Regional Total	52	54	62	168
5-19	Santa Cruz	10	21	11	42
20-99		11	21	10	42
100+		21	3	18	42
	Regional Total	42	45	39	126
5-19	Cochabamba	6	10	6	22
20-99		7	10	5	22
100+		13	1	8	22
	Regional Total	26	21	19	66
	TOTAL	120	120	120	360

## Completed Interviews, Bolivia:

Emp. Size	Location	Manufacturing	Retail	Other Services	
5-19	La Paz	15	26	24	65
20-99		16	24	28	68
100+		20	5	9	34
	Regional Total	51	55	61	167
5-19	Santa Cruz	9	20	9	38
20-99		19	17	16	52
100+		14	7	15	36
	Regional Total	42	44	40	126
5-19	Cochabamba	6	10	7	23
20-99		10	10	9	29
100+		11	1	5	17
	Regional Total	27	21	21	69
	TOTAL	120	120	122	362

## Appendix F

### Local Agency team involved in the study:

Local Agency	Encuestas & Estudios
Enumerators involved:	30 enumerators
Other staff involved:	10 supervisors, 6 editors, 6 data entry, 1 IT Manager

### Sample Frame:

Characteristic of sample frame used:	Economic Establishments Census
Source:	Encuestas & Estudios
Year:	2006, updated to 2010
Year and organism who conducted the last economic census	INE 1992
Other sources for companies statistics	Data available from Chambers of Commerce, Industry, Tax Office, but none fulfilling the requirements of this study.

### Sectors included in the Sample:

Original Sectors	2, 15, 17, 18, 24, 25, 26, 27, 28, 31, 45, 50, 51, 52, 55, 60, 70
Added Sectors	62, 63, 64

**Sample:**

Comments/ problems on sectors and locations selected in the simple:	Four main cities, which contain nearly 80% of economic activity, were selected. Selected sectors have been thought for a universal sample without taking into account specific aspects of Bolivia's industry and companies.
Comments on the sample design:	The sample in Bolivia approximates a census for SMEs, which explains the non existence of replacement in case of non response.

**Fieldwork:**

Date of Fieldwork	May 4 to November 3 2010.
Problems found during fieldwork:	83 calls x 362 interviews = 30,046 calls were needed to find interviewees. To meet interviewees after arranging a date, an average of 3 to 10 visits were needed to conclude interviews – which means more than 2500 visits to selected establishments. I

**Questionnaires:**

Problems for the understanding of questions (write question number)	<ul style="list-style-type: none"><li>➤ C9 not easy to calculate.</li><li>➤ D1a1 and D1a2 not many have their product code</li><li>➤ C9, I2,I4, J7. Not easy to calculate means or amounts</li><li>➤ No manager can calculate time dedicated to taxes requirements (j2)</li><li>➤ It is not easy to calculate number of employees per questions L1,L3, L4, L5.</li><li>➤ To measure firm size we first used L1 and L6, than only L1</li><li>➤ LACL13, LACL15, LACL17 hard to calculate hours of training per year. Not clear if corresponds per employee, per total personnel, or what</li></ul>
---	--